FEBRUARY
1962
VOLUME 26
NUMBER 1

# JOURNAL OF CONSULTING PSYCHOLOGY

AMERICAN PSYCHOLOGICAL ASSOCIATION

# JOURNAL OF CONSULTING PSYCHOLOGY

EDWARD S. BORDIN, Editor
*University of Michigan*

BOYD MCCANDLESS, Associate Editor
*University of Iowa*



*Advisory Editors:* John W. Atkinson, *University of Michigan*. Frank X. Barron, *University of California, Berkeley*. John M. Butler, *University of Chicago*. Allen T. Dittmann, *National Institute of Mental Health*. Seymour Fisher, *Baylor University*. Donald W. Fiske, *University of Chicago*. Harrison G. Gough, *University of California, Berkeley*. William L. Hays, *University of Michigan*. Jerome Kagan, *Fels Research Institute*. Paul E. Meehl, *University of Minnesota*. Daniel R. Miller, *University of Michigan*. Alan K. Rosenwald, *University of Illinois*. Saul Rosenzweig, *Washington University*. Julian B. Rotter, *Ohio State University*. Nevitt Sanford, *University of California, Berkeley*. Seymour B. Sarason, *Yale University*. Robert R. Sears, *Stanford University*. Edwin S. Shneidman, *Veterans Administration, Los Angeles, California*. Read D. Tuddenham, *University of California, Berkeley*. George S. Welsh, *University of North Carolina*.


The *Journal of Consulting Psychology* is the clinical journal of the American Psychological Association. It is devoted primarily to the publication of original research relevant to psychological diagnosis, psychotherapy and counseling, personality, and the dynamics of behavior. Case studies, theoretical contributions, descriptions of clinical techniques, and discussions of the professional problems of clinical psychology also appear in its pages.

Manuscripts submitted for *regular* or *early* publication and manuscripts for Brief Reports should be sent to the Editor: Edward S. Bordin, Department of Psychology, University of Michigan; Ann Arbor, Michigan. All manuscripts should be submitted *in duplicate* to facilitate referral to the Advisory Editors.

*Style of Manuscripts.* Manuscripts must be prepared in the style described in the *Publication Manual of the American Psychological Association* (*1957 Revision*), obtainable for $1.00 from the Washington office of the APA.

*Early Publication.* Authors may secure the early publication of an acceptable article in the first issue to go to press after its acceptance by arranging for the article to be printed as extra pages, increasing the number of pages that subscribers receive. Information about charges for early publication may be obtained from the Managing Editor.

*American Documentation Institute.* To save printing costs, materials supplementary to articles may be deposited with the ADI. To obtain copies, order by Document Number as cited in the reference footnote, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. The indicated fee must be remitted in advance. Make checks payable to Chief, Photoduplication Service, Library of Congress.

*Reprints.* Authors receive fifty reprints gratis. Additional quantities may be ordered when returning proofs.

---


HELEN ORR
*Acting Production Manager*

Communications concerning subscriptions, change of address, claims for the nonreceipt of a number, advertising, and other business matters, should be sent to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D.C. Address changes must reach the subscription office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

The *Journal of Consulting Psychology* is published bimonthly in February, April, June, August, October, and December, the yearly volume comprising approximately 576 pages. The subscription per year is $10.00, foreign $10.50, single number $2.00.

Published by the American Psychological Association, Inc.
at Prince and Lemon Sts., Lancaster, Pa.
and 1333 Sixteenth St. N.W., Washington, D.C.

Second-class postage paid at Lancaster, Pa.



© 1962 by the American Psychological Association, Inc.

JOURNAL OF CONSULTING PSYCHOLOGY

February 1962 Vol. 26, No. 1

## Contents

ARTICLES

Dependency and the Reinforcement Value of a Verbal Stimulus: *Robert B. Cairns and Michael Lewis* - 1

Comment on Levi's "Orthopedic Disability as a Factor in Human-Figure Perception": *A. B. Silverstein* 9

Reply to Silverstein's "Comment": *Aurelia Levi* - - - - - - - - - - - - - - - - - - - - - 10

Reinforcement in a Therapy-like Situation through Selective Responding to Feelings or Content: *Irene E. Waskow* - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 11

Factor Analysis of Rorschach Scoring Categories for a Population of Normal Subjects: *Robert H. Geertsma* 20

The Control of Social Desirability: A Comparison of Three Different *Q* Sorts and a Check List, All Composed of the Same Items: *William S. Kogan and Wilbert E. Fordyce* - - - - - - - - - - - - 26

A Comparison and Reinterpretation of Factor Structures of the W-B, the WAIS, and the HAWIE on Aged Persons: *Ruth M. Riegel and Klaus F. Riegel* - - - - - - - - - - - - - - - - - - - - - 31

Similarity and Success in Therapeutic Dyads: *Robert C. Carson and Ralph W. Heine* - - - - - - - - 38

Sexual Symbols, Structured and Unstructured: *Ken Lessler* - - - - - - - - - - - - - - - - - 44

The Porteus Maze Test as a Predictor of Functioning Abilities of Retarded Children: *Robert A. Dentler and Bernard Mackler* - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 50

Premorbid Behavior and Prognosis in Female Schizophrenic Patients: *Amerigo Farina, Norman Garmezy, Marian Zalusky, and Joseph Becker* - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - 56

Dependency Fantasy and Group Conformity in Ulcer Patients: *Peter Weiss and Walter Emmerich* - - 61

Unrealistic Self-Enhancement in Paranoid Schizophrenics: *Philip H. Havener and Carroll E. Izard* - - 65

The Measurement of Characteristic Patterns of Ward Behavior in Chronic Schizophrenics: *Marvin Hunter, Carmi Schooler, and Herbert E. Spohn* - - - - - - - - - - - - - - - - - - -

# Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specification given below.

2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.

3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.

4. Agrees not to submit the full report to another journal of general circulation.

## Specifications

*Brief Report.* The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the author's lines, must not exceed 85 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style of the 1957 revision of the *APA Publication Manual.* Headings, tables, and references are avoided or, if essential, must be counted in the 85 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 85-line quota:

[1] An extended report of this study may be obtained without charge from John Doe (giving the author's full name and address) or for a fee from the American Documentation Institute. Order Document No. —— from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $—— for microfilm or $—— for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

*Extended report.* Because the extended report is intended for photoduplication, and is not copy to be sent to a printer, its style should differ in several ways from that of other manuscripts: (*a*) The extended report should be typed with single spacing for economy in duplication. (*b*) Tables and figures should be placed adjacent to the text which refers to them. A caption should be typed below each figure. (*c*) Footnotes should be typed at the bottom of the page on which reference is made to them. In other respects, the full report is prepared in the style specified by the *Publication Manual.*

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 1–8

# DEPENDENCY AND THE REINFORCEMENT VALUE OF A VERBAL STIMULUS

ROBERT B. CAIRNS AND MICHAEL LEWIS

*Indiana University*

Generalized reinforcers in the form of attention, approval, and affection are believed to play a role of nuclear importance in the maintenance and control of social behavior (Dollard & Miller, 1950; Skinner, 1953). Studies of verbal conditioning, in particular, have been often cited in support of the notion that simple expressions of attention-assent constitute a reinforcing state of affairs.

In view of the inconclusive results of several verbal conditioning studies (Krasner, 1958), it seems pertinent to question whether the reinforcement stimulus used in these studies is as invariably reinforcing as it is sometimes purported to be. Indeed, by no means is it clear, either empirically or theoretically, that such experimenter behavior as a headnod and verbalization of "mm-hmmm" or "that's a good one," should be equally reinforcing for all persons. On the contrary, as Child (1954) has suggested, to the extent that the reinforcement value of such events as praise and approval is learned, one might expect there to be marked differences among persons in their evaluation of such social reinforcement.

One experimental study is particularly relevant to this expectation. Mandler and Kaplan (1956) found that the reinforcing stimulus "mm-hmmm" seemed to have little effect on the subjects' production of plural nouns over successive reinforcement periods. A subsequent interview revealed that some of the subjects felt the reinforcing stimulus had a negative meaning: these subjects thought it meant they were going too fast, giving the wrong kinds of words, or otherwise misinterpreting the experimenter's instructions. When the data from the two groups were separately analyzed, it was found that the subjects who evaluated the reinforcing stimulus as a positive event showed some evidence of conditioning in the initial reinforcement period. On the other hand, those subjects who gave a negative interpretation to the stimulus tended to show a decrease in the number of plural nouns they produced during the reinforcement period.

Theoretical statements by Sears and his collaborators (Sears, Maccoby, & Levin, 1957; Sears, Whiting, Nowlis, & Sears, 1953) regarding dependency motivation would appear to provide an account for such systematic differences among individuals in their response to social reinforcement. According to this group, dependency refers to a learned motivational system, the behavioral manifestations of which are help seeking, attention seeking, and approval seeking. Following this conceptualization of dependency, it seems reasonable to expect, with Gewirtz and Baer (1958), that approval and affection would have enhanced reinforcement value for a person with strong dependency needs. Thus, if a highly dependent individual were placed in a situation where approval was contingent on his behavior, one would predict that he would be more likely than a markedly less dependent person to produce responses that would be approved.

Recent experimental evidence has provided some support for this expected relationship between dependent behavior and response to social reinforcement. Studying the effect of mild praise on a child's performance on learning lists of paired-associate items, Cairns (1959) found that children who were observed as being highly dependent in their social interactions (ready to seek help) showed significantly greater improvement in learning under praise than did children who manifested little dependent behavior. Similar findings have been reported recently by Endsley and Hartup (1960). Comparing the performance of groups of preschool children who differed in their tendency to seek praise from the teacher, these investigators found that the most dependent children persisted longer than

did the least dependent group on a simple, repetitive motor task where persistence was associated with social reinforcement.

While the foregoing studies provide support for the hypothesized relationship between dependency and social responsivity, it would seem desirable to establish whether an individual's subjective evaluation of the reinforcement stimulus is a function of varying levels of dependency motivation. The present study was designed to investigate this problem, and to explore further the relationship between dependency and verbal conditioning.

On the basis of the above considerations, the following hypotheses were developed:

1. In a verbal conditioning situation, the High Dependent subjects would give a higher positive evaluation to the reinforcement stimulus than would the Low Dependent subjects.

2. The High Dependent subjects would condition to a significantly higher level, i.e., produce more conditioned responses in the final block of reinforced trials, than would Low Dependent subjects.

3. Subjects who judge the reinforcement stimulus to be a positive event would condition to a higher level than would subjects who assign a negative rating to the stimulus.

## METHOD

### *Subjects*

A total of 60 male college freshmen were studied. From an administration of the Edwards (1953) Personal Preference Schedule (PPS) to 625 male freshmen students at the University of Pennsylvania, 60 men who represented extreme scorers on the deference and autonomy scales were selected for final study.[1] The 30 subjects who scored high on deference (> 70th percentile) and low on autonomy (< 50th percentile) were classified as High Dependent. Likewise, those 30 subjects high on autonomy (> 75th percentile) and low on deference (< 50th percentile) were considered to be Low Dependent. A previous validation study by Bernardin and Jessor (1957) indicates that this combination of scores on the PPS is significantly related to such manifest dependent behavior as help seeking and quality of performance under conditions of verbal punishment. Further indirect or construct validation of the scales is provided by the differential response of the High and Low Dependency subjects to our recruitment procedures. Each of the university freshmen who met our criteria for selection received a letter, then a follow-up postcard requesting their participation in the study. Because PPS scores were available on all potential subjects, it was possible to compare the group who appeared with the group who did not respond. Such a comparison reveals that significantly more High Dependent subjects came in for an appointment ($\chi^2 = 4.61$; $p < .05$).

[1] The authors wish to express their appreciation to Marvin Hurvich for making these PPS records available.

### *Measures of Dependency*

Two additional dependency measures, the Interpersonal Checklist, (ICL) and Behavioral Dependency (BD) test were administered to each subject. The ICL, devised by Leary and his associates (1957), is an inventory consisting of eight clusters of 16 self-referent items. For the purposes of this study, only one cluster was scored: Octant VI, consisting of adjectives and descriptive phrases that Leary classifies as "Docile-Dependent" items. The ICL dependency score was the absolute number of items that the subjects checked in Octant VI, including such responses as "likes to give in," "dependent," "very anxious to be approved of," etc.

Since both the PPS and ICL are self-description inventories, it seemed desirable to obtain a measure which more directly reflected manifest dependent behavior. For this purpose, each subject was presented with a nearly unsolvable puzzle task designed to assess his readiness to ask for assistance when confronted with an extremely difficult problem solving situation. The task is described in an earlier paper (Cairns, 1959). The BD score was a response latency measure, i.e., the length of time the subject worked on the puzzle before making a direct plea for help. If the subject did not request help within 15 minutes, the experimenter volunteered to assist. Those subjects refusing help at this point were allowed to continue an additional 3 minutes before the experimenter discontinued testing.

### *The Conditioning Task*

A procedure similar to that employed by Sarason (1958) was followed in the design of the conditioning task. The stimulus material consisted of 149 3″ × 5″ cards on which four words were typed: below "He" appeared three verbs in the third person singular form. Since 60 cards were repeated, the subjects had a total of 209 trials on the conditioning task. In the instructions the subjects were asked to construct a simple sentence for each card, using the pronoun and one of the three words that were typed below. For the experimental subjects, the experimenter initiated the reinforcement conditions on the 11th occurrence of the response of the class to be conditioned, and continued reinforcing every such response that occurred during the presentation of the remaining cards. The reinforcing stimulus consisted of a glance up at the subject and a murmured "mm-hmm." For the control subjects, the experimenter was silent during the presentation of the cards.

The verbs appearing on the cards were roughly equated in terms of frequency of occurrence, as determined by the Thorndike-Lorge (1944) frequency count. On each card, one of the three words had strong aggressive connotations, a second had high dependency connotations, and a third was neutral. The words were selected for inclusion in either the aggressive or dependent class on the basis of ratings by 40 judges (30 undergraduate and 10 graduate students). Verbs which were rarely rated as being either dependent or aggressive constituted the neutral category. It was thus possible to construct 149 triads, each of which contained an item from the three response classes.

Twenty subjects, 10 High Dependent and 10 Low Dependent, were reinforced for constructing aggressive sentences; another 20 subjects were reinforced for emitting dependent words. Of the remaining group, 10 High Dependent and 10 Low Dependent individuals served as control subjects. The subjects were randomly assigned to the experimental treatment groups.

### *Evaluation of the Verbal Stimulus*

Since every subject was aware that the experimenter said "mm-hmmm" from time to time, it was possible to obtain an evaluation of the verbal stimulus from every experimental subject. Each subject was presented with an 11-point scale, ranging from −5.0 to +5.0, with instructions to judge the pleasantness of the "mm-hmmm." The rating appeared to be an easy one for most subjects, and was usually made with little hesitation.

A male graduate student in his early twenties administered the verbal conditioning task, which was introduced under the guise of a reading test standardization. Following the conditioning period, the experimenter excused himself with the explanation that another experimenter, engaged in a different project, would like to see the subject. At that time, the second experimenter administered the BD test and ICL according to standardized instructions. Following the subject's completion of the ICL, the first experimenter returned to obtain the subjective evaluation of the verbal stimulus.

Neither the first nor the second experimenter was aware of the subject's dependency classification at the time the experiment was conducted. Assignment of the subjects to the various experimental conditions was made by a third person.

## RESULTS

### *Agreement between the Dependency Measures*

The results show that the dependency measures were moderately related to each other. When the two dependency groups, as selected by the PPS, were compared on their ICL responses, it was found that the High Dependent subjects checked significantly more Octant VI items than did the Low Dependent subjects ($t = 3.70$; $p < .001$). The product-moment correlation between the PPS deference scale and the checklist score is $r = .44$ ($p < .005$). A comparison of the performance of the two groups on the BD measure indicates that the High Dependent subjects tended to show somewhat shorter response latencies than did the Low Dependent group ($t = 1.72$; $.05 < p < .10$). It should be noted that the High Dependent group tended to have more subjects who asked for help before the end of the first 15-minute period, as well as fewer who refused assistance until it was necessary to remove the puzzle from them. A similar positive relationship was found between the ICL dependency score and the BD task, with the group of subjects who checked the fewest items in Octant VI requiring a significantly longer time before asking for help ($t = 2.26$; $p < .05$).

TABLE 1

COMPARISON OF HIGH DEPENDENT AND LOW DEPENDENT GROUPS WITH RESPECT TO EVALUATIONS OF THE VERBAL STIMULUS

| Group | Evaluation | | |
|---|---|---|---|
| | Non-positive | Positive | Total |
| High dependent | 8 | 12 | 20 |
| Low dependent | 16 | 4 | 20 |
| Total | 24 | 16 | 40 |
| | Fisher's exact $p = .02$ (two-tailed) | | |

### *Prediction I: Subjective Evaluation of the Verbal Stimulus*

The expectation that the High Dependent and Low Dependent subjects would differ in their subjective evaluation of the verbal stimulus was clearly supported by these data. The 40 experimental subjects were separated into two groups on the basis of their interpretation of "mm-hmmm"; 16 subjects gave a positive evaluation to the stimulus and 24 subjects considered the stimulus to be either negative or neutral. The resulting 2 × 2 table is presented in Table 1. The extent to which the obtained cell frequencies deviate from the

TABLE 2

MEDIAN NUMBER OF TRIALS FOR HIGH DEPENDENT AND LOW DEPENDENT GROUPS TO REACH TEN OCCURRENCES OF REINFORCED RESPONSE CLASS

| Group | Prereinforcement series | Final reinforcement series | Change score index of conditioning |
|---|---|---|---|
| High Dependent group | | | |
| Aggressive responses ($N = 10$) | 33.0 | 26.0 | +6.5 |
| Dependent responses ($N = 10$) | 23.0 | 25.0 | +0.5 |
| Total ($N = 20$) | 25.0 | 25.0 | +4.5 |
| Low Dependent group | | | |
| Aggressive responses ($N = 10$) | 29.5 | 39.0 | −5.5 |
| Dependent responses ($N = 10$) | 27.5 | 35.5 | +0.5 |
| Total ($N = 20$) | 28.5 | 38.5 | −3.5 |

TABLE 3

RESULTS OF THE COMPARISONS OF THE HIGH DEPENDENT AND LOW DEPENDENT GROUPS IN RESPONSE TO THE VERBAL STIMULUS

| Comparison groups | Mann-Whitney $U$ value |
|---|---|
| Prereinforcement series | |
| HD (dep) vs. LD (dep) | 36.0 |
| HD (agg) vs. LD (agg) | 49.0 |
| HD (dep) vs. HD (agg) | 20.0* |
| LD (dep) vs. LD (agg) | 41.5 |
| HD (dep + agg) vs. LD (dep + agg) | 166.5 |
| Final reinforcement series | |
| HD (dep) vs. LD (dep) | 36.5 |
| HD (agg) vs. LD (agg) | 18.0* |
| HD (dep) vs. HD (agg) | 49.5 |
| LD (dep) vs. LD (agg) | 36.0 |
| HD (dep + agg) vs. LD (dep + agg) | 107.0* |
| Change score index | |
| HD (dep) vs. LD (dep) | 47.0 |
| HD (agg) vs. LD (agg) | 19.5* |
| HD (dep) vs. HD (agg) | 32.0 |
| LD (dep) vs. LD (agg) | 35.5 |
| HD (dep + agg) vs. LD (dep + agg) | 140.5* |

Note.—HD High Dependent group. LD Low Dependent group. (dep) reinforcement followed dependent word. (agg) reinforcement followed aggressive word.
* $p < .05$; two-tailed.

expected values is reflected by the Fisher probability estimate based upon the total table. It should be noted that the significance appears to be due primarily to the tendency of the Low Dependent subjects to judge the stimulus as being a nonpositive event. This trend among the Low Dependent subjects to avoid a positive rating, when tested directly by the binomial expansion, proved to be substantial ($p < .05$).

A test of the relationship between the ICL measure of dependency and the subjective evaluation of the reinforcement stimulus yielded similar results. The product-moment correlation between the ICL dependency score and the subjective ratings was small, but significant in the expected direction ($r = .33$; $p < .05$). The corresponding correlation between the puzzle scores and the evaluations of the verbal stimulus was, however, nonsignificant ($r = .23$; $p > .05$).

*Prediction II: Dependency and Conditioning*

According to the second prediction, it was expected that the High Dependent group would show a significantly higher level of conditioning than would the Low Dependent group. In Table 2 are reported the median number of trials required by the various de-

pendency groups to attain 10 occurrences of the class of responses which were reinforced. The number of trials for two series are presented: (*a*) the prereinforcement series, i.e., the first 10 occurrences of the appropriate response class; (*b*) the final reinforcement series, i.e., the last 10 occurrences of the reinforced response. In addition, the change score index of conditioning is reported. Derived by subtracting, for each subject, the number of trials in the final reinforcement series from the number of trials in the prereinforcement series, the change scores represent the most accurate estimate of the effect of the reinforcement procedures.

The results of the comparisons among the dependency subgroups are shown in Table 3. In the first, nonreinforced series, the High Dependent subjects produced words connoting dependency more readily than they produced aggressive words. The corresponding difference among the Low Dependent subjects did not approach significance.

In the final reinforcement series, the clearest difference to emerge was that between the two groups reinforced for producing aggressive sentences. Whereas the High Dependent subjects tended to increase their rate of reinforced responses, the Low Dependent group showed a significant decrement below its performance in the prereinforcement series. This negative change of the Low Dependent group, as tested by the Wilcoxon matched-pairs test, proved to be statistically significant ($p < .05$). There was, however, no substantial difference between the change scores of the High and Low Dependent subgroups which were reinforced for producing words with dependency connotations. Thus, when the overall comparison between the High Dependent and Low Dependent groups is tested, irrespective of class of response reinforced, the obtained difference just fails to reach the predetermined level of significance.

Comparable trends were present in the relationship between the ICL dependency scores and amount of change observed in the conditioning situation. The product-moment correlation between the change score index of conditioning and the ICL dependency measure was small, but was in the predicted direction ($r = .30$; $p < .05$). The change score index of conditioning was, however, unrelated to performance on the BD task ($r = .11$; $p > .10$).

### *Prediction III: Conditioning as a Function of the Evaluation of the Verbal Stimulus*

According to the final prediction, it was expected that a positive relationship would be found between the subjects' evaluations of the "mm-hmmm" and the level of conditioning attained. To evaluate this hypothesis, the dependency measures were disregarded and the positive and negative evaluators were compared in terms of their verbal conditioning performance (see Table 4). While medians for the two groups did not differ significantly in the prereinforcement series, the differences were clearly significant in the final block. Similarly, a comparison of the change scores indicates that the interaction between the two groups approached significance. To explore further the magnitude of this relationship, the change scores were correlated with the subjective evaluations of the stimulus. The obtained product-moment correlation indicates that the relationship was only a moderate one ($r = .35$).

## DISCUSSION

This study represents one phase of a research program, the general purpose of which

TABLE 4

MEDIAN NUMBER OF TRIALS FOR POSITIVE AND NONPOSITIVE EVALUATORS TO REACH TEN OCCURRENCES OF REINFORCED RESPONSE CLASS

| Evaluation | Prereinforcement series | Final reinforcement series | Change score index of conditioning |
|---|---|---|---|
| Positive ($N = 16$) | 24.5 | 25.0 | +3.5 |
| Nonpositive ($N = 24$) | 30.0 | 36.5 | −4.5 |

TABLE 5

RESULTS OF THE COMPARISONS OF POSITIVE AND NONPOSITIVE EVALUATORS IN RESPONSE TO VERBAL STIMULUS

| Comparison series | Mann-Whitney *U* value | *p** |
|---|---|---|
| Prereinforcement | 146.0 | >.10 |
| Final | 90.0 | <.01 |
| Change | 129.0 | <.10 |

* Two-tailed tests.

is to investigate the motivational factors involved in determining responsivity to external evaluation. The data reported in this paper have provided, for the most part, preliminary support for the notion that an individual's response to social reinforcement is related to aspects of his dependency behavior.

Perhaps the most significant single finding of this study was the obverse response of the Low Dependent subjects to the reinforcement procedures. Not only did the Low Dependent subjects assign significantly fewer positive ratings to the stimulus, but these subjects showed a nontrivial decline in the production of the class of responses reinforced. Furthermore, this "negative conditioning" of the Low Dependent subjects appeared to account for most of the difference between the High and Low Dependent groups, i.e., the High Dependent group showed only minimal evidence of the expected verbal conditioning phenomenon.

It should be noted that, as a group, the Low Dependent subjects tended to deny any form of dependency behavior, whether on the self-report inventories or in the limited social setting of the Behavioral Dependency test. This failure to behave in a dependent fashion, even under conditions where a dependent response would seem to be the most appropriate one, suggests that, for these subjects, dependency behaviors are strongly inhibited. If this notion is correct, the finding of a decrement in conditioned responses among the Low Dependent subjects becomes more understandable. Insofar as experimenter attention and approval constitute cues for dependency, these events would be aversive stimuli for a person who attempts to inhibit dependency behavior, and presumably would lead him to avoid responses which elicit dependency cues. Such was the case with the Low Dependent subjects, who both evaluated the "mm-hmmm" to be a nonpositive stimulus and tended to avoid responses that would lead to a recurrence of that stimulus. Thus, rather than consider these subjects as simply lacking in dependency instigation, or having low dependency needs, it seems that their behavior might more appropriately be viewed as an active inhibition of dependency strivings.

Regarding the relationship between the dependency measures and the subjects' subjective evaluations, it might be argued that the High Dependent subjects tended to rate the "mm-hmmm" as being a positive event simply because they wanted to avoid any behavior that might be interpreted as criticism of the experimenter. Accordingly, a positive rating might be obtained whether or not the High Dependent subjects had subjectively interpreted the verbal stimulus to be rewarding. Although such a response set was not investigated in this study, it seems entirely possible that something akin to "yea-saying" did play a role in the ratings. However, the finding of a significant relationship between the ratings and the subjects' production of conditioned responses seems to indicate that the ratings reflected, to some extent, the subjects' actual subjective evaluation of the verbal stimulus.

While most of the experimental data presented here can be accounted for satisfactorily in the dependency framework, it should be noted that one of the findings of this study is more difficult to interpret; namely, the failure of the experimental subjects to show evidence of strong positive conditioning. Even subjects who assigned a positive evaluation to the verbal stimulus did not show a significant increment over the prereinforcement level. Previous reports (Krasner, 1958) indicate that such factors as directness of the reinforcing stimulus, personal characteristics and prestige of the experimenter, and the nature of the response reinforced influence the absolute level of conditioning which subjects attain. A review of the present design

indicates that each of these factors may have been operative in our experimental situation, e.g., young experimenter reinforcing emotionally charged words with minimal approval. In particular, the nature of the response category reinforced seemed highly relevant for the present study. Those groups reinforced for aggressive responses tended to show evidence of a conditioning phenomena, whereas there was slight conditioning among subjects in the groups reinforced for producing dependency words. In this regard, it should be noted that the judges who selected the words initially tended to show considerably greater agreement among themselves in selecting words for the aggression category than they showed in selecting dependency words.

Finally, a comment regarding the relative predictive power of the three types of dependency measures seems appropriate. On the basis of previous results (Cairns, 1959), it was expected that the behavioral measure would be a relatively effective predictor of behavior in the conditioning situation. This expectation obviously was not borne out by the results. Whereas both self-description measures were related to the subjective evaluations of the reinforcing stimulus as well as to the amount of conditioning, the Behavioral Dependency scores were significantly related to neither. A re-evaluation of the puzzle material suggests that part of the difficulty stems from scoring simply in terms of response latency. When the data were analyzed using a refined scoring scheme that takes into account both latency and the nature of the comments the subject makes when he accepts help, somewhat clearer trends in the predicted direction are found. Even here, however, the relationships between the puzzle scores and the dependent variables fail to reach an acceptable level of significance.

## SUMMARY

The purpose of this investigation was to explore the influence of dependency motivation on the effectiveness of social reinforcement. Sixty freshman males were separated into High and Low Dependent groups on the basis of their responses to the Edwards Personal Preference Schedule.

To test the general hypothesis that expressions of assent-approval would have differential reinforcement value for High and Low Dependent individuals, 20 subjects from each group were placed in a conditioning situation where a selected class of responses was reinforced. The remaining 20 subjects served as a control group. Following the conditioning procedures, subjective evaluations of the reinforcement stimulus were obtained. All subjects were administered two additional dependency measures, the Interpersonal Checklist and a behavioral sample of the subjects readiness to behave in a dependent manner. These three measures of dependency were shown to be moderatly interrelated.

According to the first prediction, the High Dependent subjects were expected to assign a higher positive value to the reinforcement stimulus than would the Low Dependent group. While both inventory measures of dependency supported this prediction, the behavior sample failed to yield significant results. Secondly, it was expected that the High Dependent subjects would show a significantly higher level of conditioning than would the Low Dependent subjects. This prediction was only partially supported. The differences that were obtained were due primarily to the sharp decrement in performance by the Low Dependent subjects rather than to evidence of conditioning in the High Dependent subjects.

The final prediction stated that the subjects' evaluation of the reinforcement stimulus would be positively related to the amount of conditioning that occurred. This expectation was clearly supported by the data with those subjects who gave negative evaluations to the reinforcement stimulus showing a decrement in the production of reinforced responses.

## REFERENCES

BERNARDIN, A. C., & JESSOR, R. A construct validation of the Edwards Personal Preference Schedule with respect to dependency. *J. consult. Psychol.*, 1957, **21**, 63–67.

CAIRNS, R. B. The influence of dependency-anxiety on the effectiveness of social reinforcers. Unpublished doctoral dissertation, Stanford University, 1959.

CHILD, I. L. Socialization. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge: Addison-Wesley, 1954. Pp. 655–692.

DOLLARD, J., & MILLER, N. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.

EDWARDS, A. L. *Edwards Personal Preference Schedule*. New York: Psychological Corporation, 1953.

ENDSLEY, R. C., & HARTUP, W. W. Dependency and performance by preschool children on a socially reinforced task. *Amer. Psychologist*, 1960, **15**, 399. (Abstract)

GEWIRTZ, J. L., & BAER, D. M. The effect of brief social deprivation on behaviors for a social reinforcer. *J. abnorm. soc. Psychol.*, 1958, **56**, 49–56.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

LEARY, T. *Interpersonal diagnosis of personality*. New York: Ronald, 1957.

MANDLER, G., & KAPLAN, W. K. Subjective evaluation and reinforcing effect of a verbal stimulus. *Science*, 1956, **124**, 582–583.

SARASON, I. G. Interrelationships among individual difference variables, behavior in psychotherapy, and verbal conditioning. *J. abnorm. soc. Psychol.*, 1958, **56**, 339–344.

SEARS, R. R., MACCOBY, E., & LEVIN, H. *Patterns of child rearing*. Evanston, Ill.: Row, Peterson, 1957.

SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, PAULINE S. Some child-rearing antecedents of aggression and dependency in young children. *Genet. psychol. Monogr.*, 1953, **47**, 135–234.

SKINNER, B. F. *Science and human behavior*. New York: Macmillan, 1953.

THORNDIKE, E. L., & LORGE, I. *The teacher's wordbook of 30,000 words*. New York: Teachers College, Columbia University, 1944.

(Received November 28, 1960)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 9

# COMMENT ON LEVI'S "ORTHOPEDIC DISABILITY AS A FACTOR IN HUMAN-FIGURE PERCEPTION"

A. B. SILVERSTEIN

*Pacific State Hospital*

On the basis of findings recently reported in this journal, Levi (1961) claimed support for a hypothesized one-to-one relationship between the physical body and the body image. It is the present writer's contention that Levi did not present sufficient data to justify her claim. The specific hypotheses of Levi's study were that "subjects with disability of the legs will be unusually sensitive to the legs in drawings of the human figure and that subjects wth disability of the arms will be unusually sensitive to the arms in drawings"; also, that a group of back disabilities "will show a greater resemblance to the characteristics of a control group of nondisabled" (p. 253).

Using an ingenious methodology, Levi found that subjects with disability of the legs were indeed significantly more sensitive than control subjects to the legs of stick figures, and that subjects with disability of the arms were significantly more sensitive than controls to the arms of the figures. Moreover, the group of back disabilities did not differ significantly from the controls in sensitivity to the bodies of the figures. These results are entirely consistent with Levi's specific hypotheses, but in themselves they do not constitute a decisive test of her general hypothesis.

Specifically, Levi failed to provide answers to the following critical questions: Were subjects with disability of one part of the body significantly more sensitive to that part of the body than to other parts? Were they significantly more sensitive than controls to that part of the body *only?* It is at least theoretically possible that, for example, subjects with a disability of the legs were no more sensitive to the legs than to the arms, or that they were significantly more sensitive than controls to the arms, as well as to the legs. If so, the results would not support the general hypothesis.

The data which Levi did present, coupled with considerations of her method, make such a possibility appear rather unlikely, but they do not completely rule it out. In claiming that her findings support the hypothesis of a one-to-one relationship between the physical body and the body image, Levi has a responsibility to take into account *all* of her data.

## REFERENCE

LEVI, AURELIA. Orthopedic disability as a factor in human-figure perception. *J. consult. Psychol.*, 1961, **25**, 253–246.

(Early publication received July 10, 1961)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 10

# REPLY TO SILVERSTEIN'S "COMMENT"

AURELIA LEVI

*Albert Einstein College of Medicine, Yeshiva University*

The distinction that Silverstein makes between my general hypothesis and my specific hypotheses is more apparent than real. If arm-disabled subjects prove to be "unusually sensitive to the arms in drawings" (Levi, 1961, p. 253), it means that they are more sensitive to arms than to other parts.

In the event that the pattern of body-part choices is of interest to readers, a complete list of mean choices for all three experimental groups is offered.

Silverstein's second question is not really different from his first. They both ask whether the whole level of perceptual sensitivity to all parts of the body is higher in disabled than in normal persons. This problem is irrelevant to my study. The forced-choice aspect of my method sacrifices information about overall level of sensitivity for the sake of clarifying differences among kinds of sensitivity.

TABLE 1

MEANS OF THREE DISABILITY GROUPS

| Group | Mean |
|---|---|
| Group A choices: | |
| Arms | 14.17 |
| Legs | 5.0 |
| Back | 7.92 |
| Group L choices: | |
| Arms | 7.61 |
| Legs | 11.23 |
| Back | 8.15 |
| Group B choices: | |
| Arms | 9.15 |
| Legs | 4.92 |
| Back | 12.92 |

## REFERENCE

LEVI, AURELIA. Orthopedic disability as a factor in human-figure perception. *J. consult. Psychol.*, 1961, **25**, 253–256.

(Early publication received September 19, 1961)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 11–19

# REINFORCEMENT IN A THERAPY-LIKE SITUATION THROUGH SELECTIVE RESPONDING TO FEELINGS OR CONTENT [1]

IRENE E. WASKOW [2]

*University of Maryland School of Medicine*

This study is concerned with a simple type of learning that may occur early in psychotherapy. The general hypothesis is that a therapist may, by selective responding, reinforce those verbalizations that he considers important. The specific hypothesis being tested is that selective therapist responding will function as a reinforcer of the feeling or content aspects of a patient's communications.

Although there are many references in the theoretical literature to learning that occurs during psychotherapy (see Shoben, 1948, 1953 for review), only two researchers have rigorously studied events in therapy within a learning framework. Murray (1956) demonstrated increases or decreases in content areas which were encouraged or discouraged by a therapist and Dittes (1957) interpreted his results as extinction of fear connected with making embarrassing sex statements in therapy. Both of these researchers obtained their data from recordings of normally conducted psychotherapy sessions. In order to relate clearly the behavior of the therapist to the learning of the patient, in the present study the therapist's behavior was manipulated and its effects on patient behavior investigated.

Manipulation of the experimenter's behavior and the study of its effects on the subject's learning have been commonly employed by experimental psychologists. Of special relevance here are studies of verbal conditioning, in which the experimenter attempts, through selective responding, to reinforce a particular class of verbal responses. Most of these studies (see Krasner, 1958, for review) have demonstrated an increase in the frequency of the reinforced response class. Many writers have compared the procedures in these studies to the events that take place in therapy, and Krasner (1955) has specifically emphasized the value of using the verbal conditioning paradigm to investigate variables in psychotherapy. The inference here is that the therapist may selectively respond to particular patient verbalizations. The experimental conditions in most studies of verbal conditioning differ radically, however, from those which obtain in psychotherapy. For example, many of these studies have utilized only simple responses of "umhmm" or "good" as reinforcing stimuli; the verbalizations reinforced have frequently been grammatical categories such as plural nouns or the choice of "I" or "we" from among six pronouns in beginning a sentence; and the experiments have generally been conducted in an experimental laboratory situation.

A few studies have approached the therapy situation more closely, by using more complex categories of reinforcing behavior, e.g., agreement with or paraphrasing of a response (Verplanck, 1955), or by reinforcing response classes more relevant to therapy, e.g., hostile responses (Binder, McConnell, & Sjoholm, 1957; Buss & Durkee, 1958), family as opposed to nonfamily memories (Quay, 1959), and self-references (Adams & Hoffman, 1960; Rogers, 1960). Several experiments reported since this study was undertaken have specifically dealt with the reinforcement of "feeling" responses. Levine (1958) attempted to reinforce expressions of feeling in fairly free interview situations, and Salzinger and Pisoni (1958, 1960) conditioned affect responses in both normals and schizophrenics in a situation

[1] This paper is based on a PhD dissertation at the University of Wisconsin. The author is indebted to Barclay Martin under whose supervision the research was conducted, to David Grant and Carl Rogers who served on the research committee, and to Vilma Ginzberg and Tom Tomlinson who served as raters.

[2] Formerly a Postdoctoral Research Fellow at the National Institute of Mental Health, Bethesda, Maryland.

closely resembling a traditional diagnostic interview.

In the present study, the verbal conditioning paradigm was used in a laboratory situation as closely as possible resembling psychotherapy. Accordingly, the experimenter introduced the study to student volunteers as a therapy-like experience. The class of reinforcing stimuli was reflection of certain aspects of the subjects' communications, a technique utilized at times by many therapists. It was hypothesized that such reflection would fulfill Skinner's (1953) criteria for a reinforcing stimulus. The response categories reinforced, the feeling as opposed to the content aspects of subjects' communications, have been emphasized in both theoretical and research writings of client centered orientation (Rogers, 1942; Seeman, 1954), and are of relevance to the widely held view that expression of previously punished feelings in the presence of a nonpunishing therapist is essential for new learning to occur in therapy (see, e.g., Dollard & Miller, 1950).

## METHOD

The therapist selectively responded to different types of material in each of three experimental groups. In one group (Group F), she responded to references to the subjects' feelings and attitudes, in another (Group C), to the descriptive and intellectual content of the communication, and in a third group (Group FC), to a combination of feelings and content.

### *Subjects*

Thirty-six subjects were used, 12 in each of the experimental conditions. The experimenter asked students in discussion sections of an introductory psychology course to serve in the experiment for points which might improve their grades. They were told that the experiment involved research in psychological interviewing and psychotherapy and they were asked to volunteer only if they were interested in and felt that they could profit from a therapy-like experience. This procedure was used in order to obtain subjects at least minimally motivated for psychotherapy. Of the volunteers only those who had scores in the upper two-thirds of a scale of manifest anxiety (Heineman, 1953) were asked to come for preliminary interviews, since topics they would deal with might more closely resemble therapy material. A volunteer who was currently in psychotherapy was dropped.

### *Preliminary Interview* [3]

In the preliminary interview, lasting 15 minutes, the experimenter first assured subjects of the confidentiality of the tape recordings and then instructed them to:

> . . . pretend that you are having your first visit at a counseling center and . . . discuss things that you think you might talk about if you had gone to such a center for help, telling something about yourself and about matters that you have been thinking about or that have been troubling you, and so on.

Subjects were asked to talk into the microphone, and the experimenter then left the room, so that there could be no accidental differential reinforcement during this preliminary interview.

The experimenter later listened to the taped interviews and assigned each a global rating on a seven-point scale, in which a rating of 1 was given to interviews with major emphasis on the feelings of the subject, 4 to interviews with equal emphasis on feelings and content, and 7 to interviews with major emphasis on content. The position of the subjects was largely toward the content end of the scale. The 36 subjects were ranked according to these ratings and then assigned to twelve triplets, ranks 1–3 in one triplet, ranks 4–6 in the next, etc. Subjects within each triplet were randomly assigned to the three conditions, with the restriction that each group contain 10 of the 30 female subjects and two of the six males. With subjects in the three groups matched for initial level of expressing feeling and content, any differences among the groups due to differential reinforcement might be expected to emerge clearly.

### *Experimental Sessions*

Each subject was seen for four experimental interviews, two a week for 2 weeks. These sessions extended over a 6-week period, with four matched triplets seen during each consecutive 2 weeks. The experimenter functioned as therapist and conducted all of the 30-minute interviews. All interviews were tape recorded. At the beginning of the first interview, subjects were told to consider this their first session and to talk about the same types of things they were asked to discuss during the preliminary interview, the only difference being the presence of the therapist. If possible, no further instructions were given throughout the interviews. At the beginning of each, the therapist said she would like to continue as before. Occasionally, during an interview, a subject could not proceed. Only general suggestions were made at such times, such as "Any

[3] GSRs were recorded for all interviews, but will not be discussed in this article. Although the therapist's impression was that subjects were generally not very concerned about the GSR recordings, there is no data to support this impression and it is possible that they may have influenced the results.

other things you've been thinking about?" and "Maybe something that's troubled you in the past?" Such suggestions occurred only once or twice for most subjects.

The therapist attempted to respond whenever a subject paused at the end of a sentence or group of sentences concerning one topic. The chief technique used was a reflection of the subject's communication: the therapist attempted to mirror, in her own words, some aspect of what the subject had said. This type of responding, in both F and FC groups, is similar to what has been called "reflection of feeling" in the client centered literature (Bergman, 1951). In the FC group, the therapist's response included content as well as feelings, while in the F group only feelings were reflected. Although the therapist's reflections were largely based on the subjects' explicit verbal statements, she also responded to feelings indirectly expressed through voice tone and through facial and bodily movements. The therapist's responses to subjects in the C condition were similar to what has been called "restatement of content" (Snyder, 1945).

Following is an illustration of the way in which the therapist might have responded to the same subject communication in each of the three conditions.

> Subject: Sometimes my father yells at my mother—over just about nothing. But, boy, can he get mad. And I don't like it one bit when he does that.
>
> Therapist:
>
> F: You really get upset about that.
>
> FC: You feel real upset when your father gets mad at your mother about things that don't matter.
>
> C: Sometimes your father gets real mad and yells at your mother about little things.

Aside from the manipulation of the variable being studied, the therapist attempted to keep conditions as similar as possible for all three experimental groups. Since examination of early pilot interviews had suggested that the therapist sounded less interested and less warm in the C condition than in the others, she continued pilot work until it appeared that this variable was held constant. The therapist also attempted to maintain a constant frequency of responding in the three conditions, and an analysis of variance of the number of responses in each interview revealed no significant differences among the three groups. According to the therapist's retrospective account, the FC condition seemed most similar to psychotherapy and was quite easily implemented. Responses in the C condition often seemed superficial and those in the F condition sometimes seemed to distort the subject's meaning due to the separation of feelings from the context in which they appeared.

At the end of the fourth interview, subjects were given a questionnaire, which they were asked to return to the mailbox of "another psychologist interested in subjects' reactions to the experiment." The major purpose of the questionnaire was to assess, in various ways, the subjects' awareness of the selective responding by the therapist. Subjects were asked, e.g., which of several areas they thought should be stressed in therapy, what areas had actually been stressed in their sessions, etc.

Finally, the experimenter held a 5-minute interview with each subject, in which the subject discussed his hypotheses about the purpose of the experiment.

### *Rating of Interview Material*

Ratings of the subjects' responses were made directly from the tape recordings. A subject response was defined as any communication which was surrounded by two therapist responses. Brief interruptions during a therapist's response were not rated. A list of rules for classifying responses, which had been developed in pilot work, was used for guiding decisions in difficult instances. Each response was classified as predominantly F, FC, or C or, when necessary, one half-credit was given to F and FC or to C and FC. These ratings sometimes depended on more than the simple presence or absence of words denoting feeling, since tone of voice was also taken into consideration. No inferences were made, however, about underlying emotional states, and interpretation of results must be limited to the conditioning of explicit expressions of feeling.

The therapist rated all interviews, listening to an entire tape and classifying each response as it was completed. Parts of an interview were replayed only if they could not be heard clearly or if listening was interrupted. A second rater, an advanced graduate student in clinical psychology, rated 24 of the interviews for an estimate of interrater reliability. The 24 interviews were randomly selected except for the restrictions that they include two each of the first, second, third, and fourth interviews in each of the three conditions and that there be not more than one interview per subject. As a check on the objectivity of the ratings, the second rater heard only the subjects' statements and was not aware of the experimental condition in which a subject was placed. The interrater reliability of the F-FC-C profiles for the 24 interviews, as estimated by Mosier's formula (Guilford, 1954), was .81.

For testing the consistency of the therapist's responding in the three conditions, 36 10-minute segments were extracted from the interviews. These included one segment of an interview with each of the 36 subjects, and each condition was represented by a segment from the first, second, and third 10-minute periods of each of the four interviews. The therapist's responses were categorized by a graduate student, who had previously practiced making such categorizations with the therapist, as focusing predominantly on feeling (F), content (C), or feeling combined with content (FC), and the percentages of F, FC, and C responses were computed for each segment. For the entire sample of 12 segments in the F condition, the therapist made 77% F, 22% FC, and 1% C responses. In the FC condition, she made 84% FC, 14% F, and 2% C responses, and in the C

TABLE 1

ANALYSIS OF PERCENTAGE OF RESPONSES IN F, FC, AND C CATEGORIES FOR F, FC, AND C GROUPS OVER INTERVIEWS

| Source of variation[a] | *df* | *MS* | *F* |
|---|---|---|---|
| Between means | | | |
| Categories | 2 | 12.3276 | 222.52* |
| Groups × categories | 4 | .6621 | 11.95* |
| Between individual means | 66 | .0554 | |
| Between linear trends | | | |
| Categories | 2 | .0285 | 1.86 |
| Groups × categories | 4 | .0296 | 1.93 |
| Between individual linear trends | 66 | .0153 | |
| Between quadratic trends | | | |
| Categories | 2 | .0313 | 3.19 |
| Groups × categories | 4 | .0077 | — |
| Between individual quadratic trends | 66 | .0098 | |
| Between cubic trends | | | |
| Categories | 2 | .0023 | — |
| Groups × categories | 4 | .0222 | 2.06 |
| Between individual cubic trends | 64[b] | .0108 | |
| Total | 286[c] | | |

[a] No differences possible among treatment groups, since F, FC, and C scores must sum to 100% in all interviews for all groups.
[b] Degrees of freedom reduced by two, because missing values estimated for one interview (Cochran & Cox, 1950).
[c] Requirement that three scores sum to 100% results in loss of one degree of freedom for each interview, thus total *df* reduced by one-third.
* Significant at .01 level.

condition, 88% C and 12% FC responses. In 34 of the 36 samples, the therapist responded most frequently with the response called for by the experimental condition. In one FC sample, she responded predominantly with F responses, and in one F sample, with FC responses.

## RESULTS

In order to test for changes in the F-FC-C profiles of subjects over interviews, the analysis of variance applied was a trend test (Grant, 1956) analyzed so as to reveal differences in profiles (Block, Levine, & McNemar, 1951). Since the data departed to some degree from the distribution characteristics of normality and homogeneity of variance, a probability level of .01 was stipulated for accepting a result as significant. The results of the analysis are presented in Table 1 and the mean percentages for the three groups over interviews are plotted in Figure 1.

The significant interaction between categories and groups tends to support the major hypothesis: the subjects in the different experimental conditions did produce different patterns of responses in the F, FC, and C categories. Three Duncan range tests (Duncan, 1955) were used to compare the means for the groups in the three categories. Although the groups did not differ significantly in the percentage of F responses they made, Group C made a significantly higher percentage of C responses and a significantly lower percentage of FC responses than did either of the other groups. There were no significant differences between groups FC and F and no significant trends over interviews. Thus, although Group C had a greater relative amount of C throughout the experiment, there were no consistent changes in the percentages in different categories over the four interviews. The only other significant finding was the higher overall percentage of FC responses. Even in Group C, FC responses continued throughout the interviews to constitute more than 50% of the total responses.

Since differences in the percentage of C responses appeared as early as the first interview, an analysis was performed to see

whether an increase in C responses in the C group could be found within the first interview. The interview was divided into sixths according to the number of responses made and a trend analysis identical to the one cited above was performed. The results of this analysis are presented in Table 2 and the mean percentages of F, FC, and C responses for sixths of the first interview are plotted in Figure 2. Within the first interview, significant results were again found for the difference between categories and for the category-group interaction. In addition, the component of the difference between quadratic trends which may be attributed to the group-category interaction was found to be significant. Thus, although the C responses made

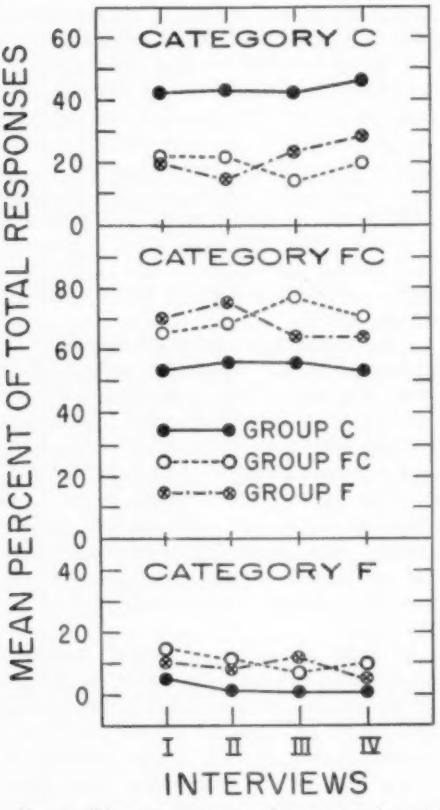


FIG. 1. Mean percentages of responses in categories F, FC, and C for three groups over interviews.

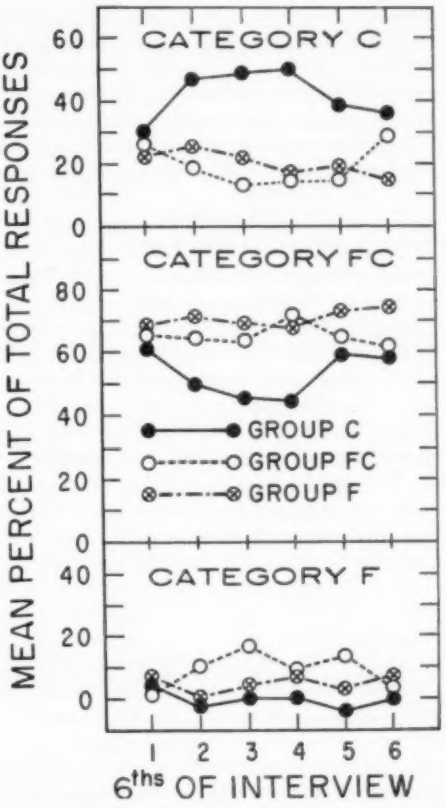


FIG. 2. Mean percentages of responses in categories F, FC, and C for the three groups during the first interview.

by the C group did not increase in a linear fashion, they did show an increase during the first part of the first interview, with a reciprocal decrease in FC responses, and it is the initial increase which seems responsible for the higher level of C responses found in the first interview for the C group. Examination of the graphs plotting sixths of the other interviews suggests that the percentage of C responses for the three groups remains fairly stable within each of the interviews following the first one.

There appeared to be no general awareness of the material being reinforced. Subjects in the three groups did not differ significantly (Friedman's test in Siegel, 1956), in the extent to which they believed different types of

TABLE 2

ANALYSIS OF PERCENTAGE OF RESPONSES IN F, FC, AND C CATEGORIES FOR F, FC, AND C GROUPS OVER SIXTHS OF THE FIRST INTERVIEW

| Source of variation[a] | *df* | *MS* | *F* |
|---|---|---|---|
| Between means | | | |
| Categories | 2 | 16.3892 | 97.32* |
| Groups × categories | 4 | .9104 | 5.41* |
| Between individual means | 66 | .1684 | |
| Between linear trends | | | |
| Categories | 2 | .0231 | — |
| Groups × categories | 4 | .0223 | — |
| Between individual linear trends | 66 | .0445 | |
| Between quadratic trends | | | |
| Categories | 2 | .0377 | 1.46 |
| Groups × categories | 4 | .2059 | 7.95* |
| Between individual quadratic trends | 66 | .0259 | |
| Between cubic trends | | | |
| Categories | 2 | .0355 | 1.35 |
| Groups × categories | 4 | .0160 | — |
| Between individual cubic trends | 66 | .0262 | |
| Residual | 144 | | |
| Total | 432[b] | | |

[a] No differences possible among treatment groups, since F, FC, and C scores must sum to 100% in all interviews for all groups.
[b] Requirement that three scores sum to 100% results in loss of one degree of freedom for each interview, thus total *df* reduced by one-third.
* Significant at .01 level.

material had been stressed in the interviews. Also, no subject was able, when questioned at the conclusion of the experiment and after administration of the questionnaire, to state correctly the purpose of the experiment or even to approach an adequate explanation.

## DISCUSSION

The major hypothesis of this study was partially supported. Selective responding to the content aspects of subjects' communications did result in a higher percentage of content responses. An increase in C responses for the C group took place within the first third of the first interview and the higher level was maintained throughout the rest of the interviews. Thus, reflection of content did function as a positive reinforcer. This finding demonstrates the occurrence of verbal conditioning in a situation closely resembling psychotherapy, in that the experiment was described to subjects as a therapy-like experience and the response class reinforced and the class of reinforcing stimuli were relevant to psychotherapy.

The successful reinforcement of content responses within the first interview leads to an interesting speculation: may such reinforcement early in therapy by some information-oriented therapists cause difficulties in their later attempts to focus on patients' feelings? To help answer this question, it would be necessary to obtain extinction as well as learning data, preferably comparing extinction of C responses through nonreinforcement and through reinforcement of F or FC responses.

Although the C group showed successful conditioning, selective responding to feelings (F group) and to a combination of feelings and content (FC group) did not function as reinforcers. For the FC group, examination of the data of the second and third interviews suggests that possibly a slower and less striking conditioning occurred than for the C group, an effect not large enough to reach

statistical significance. Since all groups had an initially high level of FC responses (between 60 and 70%), there may not have been much room for a significant increase in these responses to occur. Although the high level of FC responses throughout may seem to indicate willingness of subjects to deal with emotional material, the FC responses made were largely concerned with feelings about relatively conflict-free, conversational subject matter. The FC responses so clearly dominated the interviews that, even in the successfully conditioned C group, the increased percentage of C responses did not exceed the percentage of FC responses.

The F responses, in contrast to the FC responses, frequently dealt with personal and troubling feelings. Although the percentage of F responses was relatively low at the beginning of the experiment, thus allowing considerable room for increase, selective responding to feelings did not result in the predicted increase in these responses. Even those subjects with a relatively high initial level of expression of feelings were not successfully conditioned. In the only other study in which reflections of feeling were used in an attempt to reinforce expressions of feeling (Levine, 1958), there was a similar lack of reinforcing effects. Since Levine also failed to condition expressions of feeling with the use of an "umhmm" reinforcer, one might suspect a difficulty inherent in the class of responses being reinforced. Salzinger and Pisoni (1958, 1960), however, present contradictory evidence: they have successfully reinforced statements of affect by the use of such reinforcing stimuli as "umhmm," "yeah," and "I see." Differences between the Salzinger and Pisoni studies and the Levine "umhmm" condition will not be discussed here, since we are primarily concerned with the disparity between our results and those of Salzinger and Pisoni. There are several ways in which the present experiment differs from those of Salzinger and Pisoni. The responses reinforced in the Salzinger and Pisoni studies were defined as all statements beginning with "I" and "we" followed by an expression of affect. Thus, since the class of responses reinforced as well as the class of reinforcing stimuli was less complex and more clearly defined than that in the present study, the contingency between the reinforcing stimulus and the reinforced responses may have appeared more clearly and may, therefore, have been more effective. The response-reinforcement contingency in the present setting was, however, sufficiently clear for the reinforcement of C responses to occur.[4]

A more important difference between this study and the Salzinger and Pisoni experiments may lie in the nature of the reinforcing stimulus. The utterance of "umhmm" or "yeah" may have simple positive connotations for a subject and may thus result in an automatic reinforcement of the preceding subject verbalizations. The reflection of a feeling, however, may cause a subject to look at his feeling and to realize that another person is also aware of it. Within the context of the present experimental situation, this procedure may have been somewhat unpleasant and anxiety arousing, especially since the student volunteers seemed to prefer chatting to dealing with troubling emotions. Thus, aversive qualities of the reflection of feeling may have interfered with its proposed function as a positive reinforcer.

Under different conditions, reflection of feelings might function as a reinforcer of expressions of feelings. Recent studies (Mandler & Kaplan, 1956; Sapolsky, 1960; Sarason, 1958) have demonstrated the effects on verbal conditioning of such variables as subject personality, characteristics of the subject-experimenter interaction, and subjects' interpretation of the reinforcing stimulus. Manipulation of such variables within the context of the present study might lead to results different from those obtained. For example, individuals more similar to actual psychotherapy patients may have higher levels of expressing troubling feelings and more motivation to deal with them; they might, therefore, interpret the reflection of feelings as a positive stimulus. Also, as is probably the case in therapy, the therapist may first have to

[4] It is possible that the F condition differed from both the C condition and from Salzinger and Pisoni's procedure in having a longer latency of the therapist's responses. Although the therapist's impression is that the latencies were fairly constant in the different conditions, the recordings are no longer available for study of this variable.

establish himself as a generally positive figure before he can successfully reinforce expressions of feeling. It is interesting to note here that Salzinger and Pisoni did have a 10-minute period of introductory questions in which the therapist reinforced everything the patient said.

In the present setting, selective responding to feelings did not function as a positive reinforcer. Thus, conditioning of subjects' verbal behavior in a therapy-like setting was not as easily achieved as one might predict on the basis of the verbal conditioning literature. In order to justify generalizations to psychotherapy from the learning literature, the complex variables which may influence learning in the therapy situation must be recognized and studied.

The finding that subjects were not aware of the selective responding to certain aspects of their communications is in agreement with the frequent reports of lack of awareness in the verbal conditioning literature (see Krasner, 1958). The subjects in the present study seemed simply to accept the therapist's conduct as prescribed by a particular type of psychotherapy and were chiefly concerned with what they could find to talk about.

## SUMMARY

This study tested the hypothesis that selective responding by a therapist would function as a reinforcer of content or feeling aspects of a subject's communications. Three experimental groups were utilized, one in which the therapist reflected feeling aspects of the subjects' communications (Group F), one in which she reflected a combination of feelings and specific content (Group FC), and a third in which she reflected content only (Group C).

The 36 student subjects had volunteered to participate in a therapy-like experiment. Triplets of subjects were matched for level of expressing feeling and content in a preliminary interview and were then randomly assigned to the three groups. Each subject was seen for four 30-minute experimental interviews spaced over a period of 2 weeks. An analysis of the therapist's responses indicated that the experimental conditions in the three groups had been established reliably.

The subjects' responses were categorized as consisting predominantly of feeling elements (F), content (C), or a combination (FC). An analysis of the ratings revealed a significant difference in the F-FC-C profiles for the three groups, due to a higher percentage of C responses, and a correspondingly lower percentage of FC responses, for the C group. No significant trends were found over interviews. In a separate analysis of the first interview, a significant difference was found in quadratic trends over sixths of the interview for the group-category interaction. An initial increase in the C responses for the C group appeared responsible for the overall differences found. Thus the C group was successfully reinforced and the major effects of the reinforcement could be seen within the first third of the first interview. No significant differences were found between the F and FC groups and they showed no clear effects of reinforcement.

The questionnaire results and discussion with the subjects following the experiment suggested a lack of awareness by the subjects of the selective responding by the therapist.

The results for the content group constitute an extension of the verbal conditioning literature by demonstrating successful conditioning in a situation closely resembling psychotherapy. The lack of reinforcement in the F group suggests the presence in this setting of aversive properties in the reflection of feelings. The complex variables of the therapy situation must be considered and approached experimentally before generalizations are made from the verbal conditioning literature to psychotherapy.

## REFERENCES

ADAMS, J. S., & HOFFMAN, B. The frequency of self-reference statements as a function of generalized reinforcement. *J. abnorm. soc. Psychol.*, 1960, **50**, 384–389.

BERGMAN, D. V. Counseling method and client responses. *J. consult. Psychol.*, 1951, **15**, 216–224.

BINDER, A., MCCONNELL, D., & SJOHOLM, NANCY A. Verbal conditioning as a function of experimenter characteristics. *J. abnorm. soc. Psychol.*, 1957, **55**, 309–314.

BLOCK, J., LEVINE, L., & MCNEMAR, Q. Testing for the existence of psychometric patterns. *J. abnorm. soc. Psychol.*, 1951, **46**, 356–359.

BUSS, A. H., & DURKEE, ANN. Conditioning of hostile verbalizations in a situation resembling a clinical interview. *J. consult. Psychol.*, 1958, **22**, 415–418.

COCHRAN, W. G., & COX, GERTRUDE M. *Experimental designs*. New York: Wiley, 1950.

DITTES, J. E. Extinction during psychotherapy of GSR accompanying "embarrassing" statements. *J. abnorm. soc. Psychol.*, 1957, **54**, 187–191.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.

DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

GUILFORD, J. P. *Psychometric methods*. (2nd ed.) New York: McGraw-Hill, 1954.

HEINEMAN, C. E. A forced-choice form of the Taylor anxiety scale. *J. consult. Psychol.*, 1953, **17**, 447–454.

KRASNER, L. The use of generalized reinforcers in psychotherapy research. *Psychol. Rep.*, 1955, **1**, 19–25.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

LEVINE, G. The effects of two verbal techniques on the expression of feelings. Unpublished doctoral dissertation, Columbia University, 1958.

MANDLER, G., & KAPLAN, W. K. Subjective evaluation and reinforcing effect of a verbal stimulus. *Science*, 1956, **124**, 582–583.

MURRAY, E. J. A content-analysis method for studying psychotherapy. *Psychol. Monogr.*, 1956, **70** (13, Whole No. 420).

QUAY, H. The effect of verbal reinforcement on the recall of early memories. *J. abnorm. soc. Psychol.*, 1959, **59**, 254–257.

ROGERS, C. R. *Counseling and psychotherapy*. Boston: Houghton Mifflin, 1942.

ROGERS, J. M. Operant conditioning in a quasi-therapy setting. *J. abnorm. soc. Psychol.*, 1960, **60**, 247–252.

SALZINGER, K., & PISONI, STEPHANIE. Reinforcement of affect responses of schizophrenics during the clinical interview. *J. abnorm. soc. Psychol.*, 1958, **57**, 84–90.

SALZINGER, K., & PISONI, STEPHANIE. Reinforcement of verbal affect responses of normal subjects during the interview. *J. abnorm. soc. Psychol.*, 1960, **60**, 127–130.

SAPOLSKY, A. Effect of interpersonal relationships upon verbal conditioning. *J. abnorm. soc. Psychol.*, 1960, **60**, 241–246.

SARASON, I. G. Interrelationships among individual difference variables, behavior in psychotherapy, and verbal conditioning. *J. abnorm. soc. Psychol.*, 1958, **56**, 339–344.

SEEMAN, J. Counselor judgments of therapeutic process and outcome. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954.

SHOBEN, E. J. A learning-theory interpretation of psychotherapy. *Harv. educ. Rev.*, 1948, **18**, 129–145.

SHOBEN, E. J. Some observations on psychotherapy and the learning process. In O. H. Mowrer (Ed.), *Psychotherapy theory and research*. New York: Ronald, 1953.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

SKINNER, B. F. *Science and human behavior*. New York: Macmillan, 1953.

SNYDER, W. U. Investigation of the nature of nondirective psychotherapy. *J. gen. Psychol.*, 1945, **33**, 193–223.

VERPLANCK, W. S. The control of the content of conversation: Reinforcement of statements of opinion. *J. abnorm. soc. Psychol.*, 1955, **51**, 668–676.

(Received October 26, 1960)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 20–25

# FACTOR ANALYSIS OF RORSCHACH SCORING CATEGORIES FOR A POPULATION OF NORMAL SUBJECTS

ROBERT H. GEERTSMA

*Neuropsychiatric Institute, University of California, Los Angeles*

The use of a normal rather than a patient population of subjects for factor analytic studies of the Rorschach offers the opportunity to check hypotheses about dimensions of the test without complicating the analysis by the possible effects of emotional disturbance. Despite this significant advantage, only two factor analytic studies to date have been based on normal populations. Their results and conclusions, although not entirely consistent, are of considerable interest. Wittenborn (1950) studied the Rorschach responses of 92 Yale undergraduates. Using as variables 20 response categories representing the main scoring categories of the Klopfer system, he extracted four centroid factors from his correlation matrix. After rotating these factors to an orthogonal simple structure solution he concluded:

1. The factorial composition of *M* differs from that of *CF* and *C*. The factorial composition of *CF* and *C* is similar.
2. The factorial composition of *FC* is different from that of *CF* and *C*. It is more similar to *M* than to *CF* and *C*.
3. There is a factor loading *M* (and various detail categories) which is different from factors loading *W*, *CF*, *C*, texture, and diffusion.
4. The various scoring categories correlate with *R* to produce a general productivity factor; this factor accounts for differences in the common factor variance among the scoring categories.
5. The various subclasses of scoring categories (e.g., *FC*, *CF*, and *C*) within the broader classes (e.g., color, texture, shading) have dissimilar factorial compositions which could not be predicted from current interpretative principles.

In a recent study, Wishner (1959) presented a factor analysis of Rorschach scoring categories taken from data gathered by Beck, Rabin, Thiesen, Molish, and Thetford (1950) from a stratified population of 157 normal subjects. Wishner advanced the following hypotheses to guide his investigation:

1. A factor of psychological efficiency can be found.
2. The factorial composition of *FC* is different from that of *CF* and *C*.
3. A factor of verbal productivity can be found.

From his 20 variable (representing Beck's main scoring categories) correlation matrix Wishner extracted four factors by the centroid method. After failing to rotate these factors to a satisfactory simple structure solution, he used his hypotheses and the data supporting them to guide his placement of hyperplanes in the rotational process. After evaluating the degree of agreement between his final rotational solution and his original hypotheses, he reported confirmation of Hypothesis 1, disconfirmation of Hypothesis 2, and support for Hypothesis 3. Wishner's results failed to confirm Wittenborn's finding that the factorial composition of *FC* differs from that of the other color categories, *C* and *CF*.

The present analysis was undertaken in an effort to realize Wishner's original plan of rotating factors derived from his correlation matrix to an acceptable simple structure solution. In order to eliminate the possibility that Wishner's failure to achieve a simple structure solution could be due to the underextraction of factors, his correlation matrix was refactored by the principal components method and seven first-order factors extracted.[1] The

[1] Following the suggestion of Kaiser (1960), factor extraction was continued until the latent roots became less than 1.00. The seventh factor has a root less than 1.00 but was included in the rotational work in order to insure a sufficiency of factors for an adequate rotational solution.

TABLE 1

SCORING CATEGORIES, UNROTATED AND ROTATED FACTOR LOADINGS

| Scoring categories | Unrotated factor loadings | | | | | | | | Rotated factor loadings | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | I | II | III | IV | V | VI | VII | $h^2$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| *R* | −04 | 78 | −14 | −16 | −17 | −03 | 22 | 73 | 13 | −65 | 17 | 06 | −17 | 11 | 25 |
| *W*% | 68 | −48 | 18 | 30 | −31 | −09 | 15 | 94 | 25 | 17 | 00 | −02 | 80 | 06 | −03 |
| *D*% | −46 | 21 | −10 | −51 | 50 | 28 | −22 | 90 | 04 | 20 | −02 | −07 | −92 | −09 | −02 |
| *Dd*% | −49 | 55 | −12 | 21 | −16 | −24 | 03 | 68 | −48 | −56 | 07 | 15 | −06 | 03 | 06 |
| *M*% | 15 | 40 | 81 | −13 | −09 | 06 | −03 | 88 | 05 | 08 | 92 | −07 | 08 | −03 | 04 |
| (*C* + *CF*)% | 81 | −02 | −17 | 28 | −16 | 29 | 00 | 87 | 86 | −04 | −06 | −13 | 09 | 08 | −09 |
| (*C* + *CF* + *FC*)% | 88 | −05 | −17 | −36 | −03 | 02 | 12 | 95 | 86 | −02 | −09 | 07 | 13 | −09 | 07 |
| *FC*% | 49 | −08 | −06 | −28 | 23 | −47 | 25 | 66 | 35 | 04 | −08 | 41 | 10 | −35 | 30 |
| Sum *C*/*R* | 87 | −03 | −17 | −32 | −08 | 18 | 06 | 92 | 87 | −02 | −08 | −03 | 10 | 02 | −02 |
| (*Y* + *YF* + *FY*)% | 48 | 32 | −15 | 28 | 47 | −44 | −11 | 86 | −05 | 00 | −07 | 89 | 01 | −04 | −15 |
| (*V* + *VF* + *FV*)% | 23 | 36 | −08 | 55 | 11 | 58 | 30 | 94 | 03 | −01 | −07 | 06 | 13 | 93 | 08 |
| (All *Vs* + *Ys*)% | 54 | 48 | −12 | 50 | 44 | −04 | 06 | 99 | 00 | 00 | −03 | 75 | 07 | 47 | −08 |
| *F*% | −79 | −37 | −27 | 22 | −15 | 24 | 10 | 97 | −49 | 01 | −45 | −45 | −07 | 22 | 07 |
| *F* + % | 26 | −21 | 35 | 10 | 63 | 27 | 10 | 72 | 06 | 75 | 13 | 23 | −10 | 30 | 08 |
| *H*% | −20 | 45 | 78 | −06 | −07 | −04 | −04 | 86 | −26 | 01 | 88 | −05 | 00 | −07 | 07 |
| *A*% | −39 | −62 | −10 | −14 | 26 | −09 | 40 | 80 | −16 | 38 | −42 | −14 | −05 | −10 | 47 |
| *N*/*R* | 47 | −26 | −21 | 26 | −04 | 04 | −70 | 89 | 12 | 14 | −18 | 18 | 10 | −09 | −82 |
| *Z*/*R* | 72 | −25 | 33 | 20 | −24 | 00 | 06 | 79 | 32 | 18 | 24 | 00 | 64 | 07 | −09 |
| *P*% | −08 | −65 | 41 | −04 | 32 | 03 | −09 | 71 | −14 | 76 | 09 | −08 | 00 | −16 | −03 |
| *T*/*R* | −30 | −52 | 14 | 29 | −03 | −17 | 06 | 50 | −45 | 27 | −12 | −07 | 31 | −06 | 04 |

Note.—Decimal points are omitted.

use of more than four factors was recommended by our interest in identifying as many as feasible of the independent dimensions underlying the scoring categories used as variables. Most clinicians would probably expect these categories to represent more than four dimensions. The seven factors were then rotated analytically by Carroll's (1953; Kaiser, 1958) oblimin method to an oblique simple structure solution.[2] Using the same computational procedures, two second-order factors were obtained and rotated toward simple structure.

## RESULTS

The findings of this study are based on the correlation matrix presented by Wishner (1959, Table 2).

The data collection procedures and the population of normal subjects are described by Beck et al. (1950), and by Wishner (1959). Of the 157 subjects, 71 were males and 86 were females. Their ages ranged from 17 to 69, and their occupations were divided among executives, junior executives, skilled workers, semiskilled workers, and unskilled workers.

Table 1 contains the scoring categories, original factor loadings, and rotated first-order factor loadings (reported in the reference vector system). The analytically rotated factor loadings can be seen to represent a compelling simple structure solution, with the possible exception of the first factor. Although further rotation of Factor 1 seems desirable to place more loadings in the near-zero range, this was not attempted because the interpretation of this factor proved to be clear as it stands. Table 2 presents the correlations between the primary factors. The calculation of second-order factors, given in Table 3, was undertaken in order to illuminate the major dimensions underlying the scoring categories. The above-stated hypothe-

TABLE 2

CORRELATIONS BETWEEN PRIMARY FACTORS

| Factors | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | 1.00 | | | | | | |
| 2 | .05 | 1.00 | | | | | |
| 3 | .09 | −.13 | 1.00 | | | | |
| 4 | .32 | −.07 | .13 | 1.00 | | | |
| 5 | .16 | .18 | −.12 | .03 | 1.00 | | |
| 6 | .08 | −.16 | .19 | .13 | −.11 | 1.00 | |
| 7 | −.12 | .04 | .03 | .04 | −.13 | −.13 | 1.00 |

[2] Thanks are due to John Carroll for providing his IBM 704 Oblimin rotational program, and to the Western Data Processing Center, U.C.L.A. Graduate School of Business Administration, for use of their computer facilities.

TABLE 3

SECOND-ORDER FACTOR LOADINGS

| Factor | Rotated loadings | |
|---|---|---|
| | *S1* | *S2* |
| 1. Color | 18 | 79 |
| 2. Perceptual accuracy | 55 | −02 |
| 3. Human movement | −51 | 31 |
| 4. Shading | −15 | 72 |
| 5. Organizational ability | 64 | 31 |
| 6. Vista | −58 | 28 |
| 7. Restricted content | −37 | −19 |

Note.—Decimal points are omitted.

ses about psychological efficiency and verbal productivity factors might possibly be confirmed in the second-order rather than the first-order realm.

## INTERPRETATION OF THE FACTORS

Let us next take up the meaning of the seven first-order factors in terms of the variables loading high on them. For this purpose all loadings outside of the limits ±.31 (i.e., 10% of the variance) will be considered to be significant.

*Factor 1.* This factor has high positive loadings on Sum $C/R$, $(C + CF)\%$, and $(C + CF + FC)\%$, $Z/R$, and $FC\%$, and negative loadings on $F\%$, $Dd\%$, and $T/R$. This appears to be a *color factor* and to include the *FC* as well as *C* and *CF* categories. The significant $Z/R$ loading suggests that this color factor is associated with combinational ability, an association which implies that some flexibility and intellectual skills are represented by *C* and *CF* responses. We can interpret the negatively loaded $F\%$ and $Dd\%$ variables as due to the circumstance that color responses cannot be *F* responses and are not apt to be *Dd* responses in normals.

*Factor 2.* The second factor has high positive loadings on $P\%$, $F + \%$, and $A\%$, and negative loadings on *R* and $Dd\%$. This is clearly a *perceptual accuracy factor.* The negative loadings on *R* and *Dd* suggest that high perceptual accuracy is associated in normals with a restriction on *R* and *Dd*.

*Factor 3.* This factor loads high on $M\%$ and $H\%$ positively, and $F\%$ and $A\%$ negatively. This is a *human movement factor.* The negative loadings can be attributed to the mutual exclusiveness of *M* and *H* with regard to *F* and *A* in the scoring system.

*Factor 4.* The high loadings on this factor are $(Y + YF + FY)\%$, (all *Y*s and *V*s)%, $FC\%$ positively, and $F\%$ negatively. This is clearly a *shading factor,* and one which is loaded *FC* but not *C* and *CF*. The high loading on (all *Y*s and *V*s)% can be attributed to the influence of the *Y* responses in this variable.

*Factor 5.* Positive high loadings on this factor are $W\%$, $Z/R$, and the negative loading is $D\%$. This is a factor or *organizational ability.* Intelligence is also probably related to this factor, but in the absence of evidence of its correlation with an external criterion of intelligence, we will call it an organizational ability factor.

*Factor 6.* This factor has high loadings on $(V + VF + FV)\%$ and (all *Y*s and *V*s)% positively, and $FC\%$ negatively. Factor 6 may be termed a *vista factor.* The high loadings on (all *Y*s and *V*s)% can be attributed to the vista responses included in the scoring of this variable.

*Factor 7.* The seventh factor has a negative loading on $N/R$ for its highest loading. $A\%$ is the high positive loading. This factor seems to refer to a *restricted range of content.*

The two second-order factors appear to represent dimensions of cognitive and affective responsivity, respectively. The cognitive dimension, factor *S1*, is bipolar. Drawing from current interpretative principles, we speculate that on one extreme the high loadings for perceptual accuracy and organizational ability suggest an external focus of cognitive activity, whereas the negative high loadings on vista, human movement, and restricted-content factors identify the opposite pole with internally focused cognitive activity. The *S2* factor with its major loadings for color and shading clearly represents the dimension of affective experience.

## DISCUSSION

Looking at the $h^2$ column of Table 1, we note that the seven first-order factors account for at least 50% of the variance for each of the 20 variables. The studies of Wit-

tenborn and Wishner, in contrast, both reported that seven of their 20 variables did not contribute as much as 50% of their variance to the results. Thus, the present findings can be considered well grounded in the common factor variance of the variables under analysis.

It is most striking that the basic scoring categories represented in the 20 variables show such minimal overlapping and intermixing in the structure of the first-order factors. The identification of color, human movement, shading, organizational ability, vista, and perceptual accuracy factors which are only slightly correlated provides extremely cogent support for the separate scoring of these variables.

The second-order factors uphold the familiar distinction between affective and cognitive dimensions in the Rorschach test. And yet, these dimensions overlap just as we expect affect and cognition to interact in the functioning personality. The shading and color factors have only negligible loadings on the cognitive factor, but the human movement, organizational ability, and vista factors all contribute some variance to the affective dimension. The implication is that *M*, *Z*, and *V* all involve some affective responsivity secondary to their cognitive meaning but that this responsivity is much weaker than that represented by the color and shading responses. In this context, the finding that perceptual accuracy has no comparable loading on the affective factor suggests that perceptual accuracy is the variable most antithetical to affective responsivity.

The structure of the second-order factors suggests that the vista variable is more closely related to a cognitive process than to affect or a feeling experience, as represented by shading or color responses. The second-order loadings (Table 3) for the vista factor (Factor 6) closely parallel the loadings for the human movement factor (Factor 3), and we interpret this as indicating: (*a*) the cognitive activity of vista responses, and (*b*) the likelihood that vista responses, like human movement responses, involve or are associated with internal representations of people. The latter point fits well the interpretation of vista responses as self-appraisal. Rorschach test users may take these findings to indicate that dysphoric affect is not necessarily implied by vista responses, and that some sensitivity to the self and others as objects of thought is implied by the vista response.

An important general question can be answered from the results of the second-order factors, namely the question of the fundamental dimensions of Rorschach's test as represented in the present variables. Four clusters of variables are distinguishable, *C*s and *Y*s, *F*+ and *P*, *M* and *V*, *W* and *Z*. The other variables, *R*, *FC*, *F*, range of content, and *T*/*R*, etc. all relate to these main lines of the personality configuration after the fashion indicated by the first-order factors.

Let us now consider Wishner's three hypotheses.

1. The hypothesis of a psychological efficiency dimension proposes a factor representing a continuum running from flexibility of response at one pole to rigidity of response at the other. Wishner regarded *F*% as most representative of the rigidity pole because of its correlation with general alertness (Wishner, 1953). He expected *T*/*R* and *A*% also to represent the rigidity pole and *M*%, *C*, *Y*, and *V* to define the opposite pole, flexibility. Factor 1 fits part of this expectation very closely, for we find the color determinants and *Z*/*R* at the positive pole (i.e., spontaneous emotional responsivity and the ability to educe relationships) and *F*%, *Dd*%, and *T*/*R* (i.e., attention to form, attention to small details, and deliberateness) on the negative side. Factor 1, then, can be interpreted as a bipolar factor representing a continuum of flexibility—rigidity. We should, however, also note that the *F*% variable loads negatively on two other factors, thus suggesting a possible flexibility—rigidity dimension in regard to human movement (Factor 3) and shading (Factor 4) as well as for Factor 1. It is in this broad sense that the present results support Wishner's first hypothesis; the color, shading, and human movement categories can each be viewed as one pole of a continuum which has attention to form as the opposite dimension. In the present findings, the *F*% variable serves more as a contrasting effect within the various factors than as representing a basic dimension in either the first or second-order factor space.

2. The factorial composition of *FC%* in the present results is both different from and more complex than that of *(C + CF)%*. Yet it is not completely different from *(C + CF)%*. Although *FC%* has a relatively low but significant loading on the *(C + CF)%* factor (Factor 1), it also loads just as highly on two other factors, the shading factor (Factor 4) positively, and the vista factor (Factor 6) negatively. The high loading of the *FC* variable on the shading factor indicates that the affective experience of *FC* contains, along with the color component, an equal measure of the shading response. Interpretatively, we may speculate that it is the shading component that modulates, inhibits, opposes, or controls the experience of the color responses. In any case, Rorschach workers have for many years thought of *FC* as being different from *C* and *CF* responses, at least in the sense that *FC* is much further from *CF* than *CF* is from *C*. Although the factorial composition of *FC%* supports Wittenborn's conclusion about differing subgroup factorial compositions existing within a class of scoring determinants, it cannot be said that this finding is not consistent with current interpretations of the meaning of the *FC* response.

3. The verbal productivity hypothesis posits a factor with high loadings on *Dd%*, *M%* and *(C + CF)%*. The present analysis, however, gives the *M%* and *(C + CF)%* variables separate factorial status, so this hypothesis is not confirmed in its explicitly stated form for either the first or second-order factors. Nevertheless, productivity seems involved in the perceptual accuracy factor (Factor 2) which is loaded negatively by *Dd%* and *R*. This factor suggests that perceptual accuracy is contrasted with a great number of responses. In the second-order realm, the positive pole of *S*1 is also associated with perceptual accuracy. Also related to Wishner's third hypothesis is the factor we have called organizational ability (Factor 5), and we note that this factor, like Factor 2, is represented on the positive pole of *S*1. Factor 5 might be interpreted as an intelligence factor such as Williams and Lawrence (1953) reported for a population of psychiatric patients. Their intelligence factor loaded *W*, *M*, *F*, *FC*, and *R* as well as the Wechsler-Bellevue Verbal and Performance IQ scores. However, Factor 5, with its *W* responses, organizational ability, and deliberateness appears more clearly defined either than the intelligence factor of Williams and Lawrence or than is implied by Wishner's statement of his third hypothesis.

Despite the above mentioned points of relevance, the present findings do not fit well into Wishner's explicit expectations. The separate status of the basic scoring categories found here militates against the emergence of a verbal productivity or a psychological efficiency factor. Wishner's Hypotheses 1 and 3 assume that the various *C*, *M*, *Y*, and *V* variables which should load on these factors are not themselves independent dimensions, and this assumption seems to be supported by his previous work (Wishner, 1953). Methodological differences were perhaps most significant in producing the contrast between Wishner's work and that reported here. The use of a relatively large number of factors distinguishes this study most clearly from Wishner's procedure. If only four factors had been used in this study, then, because some seven semiindependent dimensions are represented in the data, one would expect both a difficulty in rotating to a simple structure solution and a mixing of the seven dimensions to produce factors that combine these dimensions. The use of seven factors for this analysis seems to have avoided these problems and to have produced a different character of findings.

## SUMMARY

In factor analyzing the Rorschach scores of 157 normal subjects whose protocols were collected by Beck et al. (1950), Wishner (1959) hypothesized that: (*a*) There is a psychological efficiency factor; (*b*) The factorial composition of *FC* is different from that of *C* and *CF* combined; and (*c*) There is a verbal productivity factor. After first failing to rotate his four factors to an acceptable simple structure solution, he used his hypotheses to guide his factor rotation and to provide criteria of evaluation. In this fashion, he was able to report confirmation of the first hypothesis, no confirmation of the second, and some support for the third.

The present study reiactored Wishner's 20 variable correlation matrix, extracting seven first-order factors. These were analytically rotated to a simple structure solution and were interpreted as: (*a*) a color factor, (*b*) a perceptual accuracy factor, (*c*) a human movement factor, (*d*) a shading factor, (*e*) an organizational ability factor, (*f*) a vista factor, and (*g*) a restricted range of content factor.

The correlations between the above factors were then factor analyzed to produce two second-order factors which were also analytically rotated to a simple structure solution. These factors were interpreted as: (*S1*) a cognitive factor, and (*S2*) an affective factor. The close parallel between the first-order factors and the Rorschach scoring categories was discussed. The *FC%* variable's loading more significantly on the shading factor than on the color factor was noted, and the conclusion was drawn that the vista variable implies internal cognitive activity akin to that represented by the human movement variable.

The relationships between the above factors and Wishner's hypotheses were discussed, and it was concluded that the present results provide partial support for his first two hypotheses. It was suggested that the differences between this analysis and Wishner's study could be primarily due to the number of factors extracted and rotated.

## REFERENCES

BECK, S. J., RABIN, A. J., THIESEN, W. C., MOLISH, H., & THETFORD, W. N. The normal personality as projected in the Rorschach test. *J. Psychol.*, 1950, **30**, 241–298.

CARROLL, J. B. An analytical solution for approximating simple structure in factor analysis. *Psychometrika*, 1953, **18**, 23–38.

KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, **23**, 187–200.

KAISER, H. F. The applications of electronic computers to factor analysis. *Educ. psychol. Measmt.*, 1960, **20**, 141–151.

WILLIAMS, H. L., & LAWRENCE, J. F. Further investigation of Rorschach determinants subjected to factor analysis. *J. consult. Psychol.*, 1953, **17**, 261–264.

WISHNER, J. Neurosis and tension: An exploratory study of the relationship of physiological and Rorschach measures. *J. abnorm. soc. Psychol.*, 1953, **48**, 253–260.

WISHNER, J. Factor analyses of Rorschach scoring categories and first response times in normals. *J. consult. Psychol.*, 1959, **23**, 406–413.

WITTENBORN, J. R. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.*, 1950, **14**, 261–267.

(Received October 31, 1960)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 26–30

# THE CONTROL OF SOCIAL DESIRABILITY:

## A COMPARISON OF THREE DIFFERENT *Q* SORTS AND A CHECK LIST, ALL COMPOSED OF THE SAME ITEMS

WILLIAM S. KOGAN

*Veterans Administration Hospital, Seattle, Washington*

AND WILBERT E. FORDYCE

*University of Washington, School of Medicine*

In the field of personality inventories considerable energy in recent research has been directed toward the problem of isolating and eliminating sources of variation that are confounded with the variables an inventory is supposed to measure. This variance has usually been considered distorting and capricious in its effect unless removed. The distorting variance has at various times been called test-taking attitude, response set (Cronbach, 1946), etc. One illustration of these measures of unwanted variances is the *F, K, L* scales of the MMPI (Edwards, 1957a, 1957b; Meehl & Hathaway, 1946). The measures obtained are used as a "correction" for the other obtained scores. Edwards (1953b), defining the principal source of the unwanted variation as social desirability (SD), hypothesized and demonstrated a stable relationship between median social desirability (MSD) value of an item and the probability of its endorsement. He devised an inventory, Personal Preference Schedule (1953a), whose design attempted to resolve the effect of SD by minimizing it as a source of variation.

Concurrently with the delineation of some characteristics of these unwanted variances and methods for controlling them, there was increasing interest in the possibility that they might have value and meaning for personality assessment. For example, *F, K,* and SD were found to be highly related among themselves and with several of the clinical scales of the MMPI (Fordyce, 1956). In another study SD was shown to be related to a sickness-health continuum (Kogan, Quinn, Ax, & Ripley, 1957). Therefore, it seemed important to unconfound but retain for measurement rather than remove or minimize the variance of SD. A similar position is expressed by Broen and Wirt (1958).

This study was designed to compare four rating scale forms [1] with respect to two general questions: (*a*) their equivalence, i.e., correlations between corresponding variables on the different forms, and (*b*), the differential effectiveness of different ways of controlling for SD.

Within the question of equivalence a specific question is, do a *Q* sort and a check list give comparable results. Block (1957) has answered this in the affirmative but his data were not based upon identical items. Within the question of differential effectiveness of ways of controlling for SD, we were interested in comparing first, check list and *Q* sorts; second, within the *Q* sorts the effect of varying the degree of comparability of the MSD values of items competing for endorsement.

It is important to note that this study does not involve validation. It is a methodological study concerned with elucidating characteristics of the rating scale forms so that some practical decisions can be made as to which form is likely to produce variable scores least confounded with SD and also provide a measure of SD.

### METHOD

A number of articles (La Forge & Suczek, 1955), in the literature dating back to 1951 described the construction and application of an 128-item check list emphasizing interpersonal rather than intrapersonal variables. An interesting characteristic of this interpersonal check list was that its construction imparted some equivalence of the MSD values to its variables. It was therefore used in

[1] Copies of the forms can be obtained on request from William S. Kogan, Veterans Administration Hospital, Seattle, Washington.

the present study as a reservoir of items having both a face validity structure and some known empirical characteristics. From this reservoir of 128 items (eight variables, 16 items each) a subsample of 64 items (eight variables, 8 items each), was constructed such that the sum and range of MSD values for each of the eight variables of the instrument were quite closely equated. SD values were obtained as described by Edwards (1957a, 1957b) on an independent sample of patients. Table 1 gives the sum and range of MSD values for the items for each of the eight variables.

### *Construction and Scoring of Rating Scale Forms*

These 64 items were used in constructing the four forms studied. Forms 1 and 2 each consisted of four subsets of 16 items each, 2 from each variable. In Form 1, each subset of 16 items contained a restricted and comparable level of item SD values. Form 2, each subset had a range of item SD values, selected such that each of the eight variables had one comparably high and one comparably low SD-valued item. In Forms 1 and 2, subjects sorted each subset of items separately into a forced normal distribution. The forced normal distribution was 2, 3, 6, 3, 2. Variable scores were calculated by summing the item placements across the four subset sortings. Form 3 consisted of a single 64-item *Q* sort. The forced normal distribution was 2, 5, 7, 11, 14, 11, 7, 5, 2. Form 4 consisted of a check list form in which the items were listed on a page and subjects simply checked each item as being characteristic or uncharacteristic of the self. Table 2 defines the term "level" employed in the text with reference to Forms 1 and 2.

It will be noted that Forms 1, 2, and 3 offer, respectively, an increasing opportunity for items of differing MSD value to compete with each other. Form 4 provides an opportunity to compare the *Q* sort, or forced-choice forms, with the more common inventory form.

In addition to yielding scores on each of the eight variables, Forms 2, 3, and 4 yield SD scores. These are calculated by summing the degree of endorsement of the 32 items (four from each variable) having the higher MSD values. In Form 4, the SD score is simply the number of these 32 items endorsed. The SD score becomes a variable to be correlated with other test variables. The average intercorrelation of the SD scores on the three forms was .76. In this study the SD score obtained from Form 3 was arbitrarily selected as the one to be correlated with variable scores from all forms.

TABLE 1

VARIABLES OF INTERPERSONAL CHECK LIST AND SUM AND RANGE OF MEDIAN SD VALUES OF ITEMS IN VARIABLES

| Variable | | MSD | |
|---|---|---|---|
| Name | Code | Sum | Range |
| Managerial—Autocratic | PA | 39.28 | 2.21–7.59 |
| Competitive—Exploitive | BC | 39.67 | 2.33–7.25 |
| Blunt—Aggressive | DE | 39.07 | 2.38–7.76 |
| Skeptical—Distrustful | FG | 37.29 | 2.54–6.23 |
| Modest—Self-effacing | HI | 39.35 | 2.50–7.31 |
| Docile—Dependent | JK | 38.64 | 2.85–7.13 |
| Cooperative—Overconventional | LM | 42.60 | 2.94–7.47 |
| Responsible—Overgenerous | NO | 41.70 | 3.67–7.33 |

TABLE 2

MEAN AND RANGE MEDIAN SD OF LEVELS

| Level | Mean | Range |
|---|---|---|
| High | 7.0 | 5.6–7.8 |
| Middle high | 6.2 | 5.0–7.3 |
| Middle low | 3.6 | 2.3–4.7 |
| Low | 2.9 | 2.2–3.7 |
| All | 4.9 | 2.2–7.8 |

### *Subjects and Administration*

The subjects were 39 male veterans hospitalized on a psychiatric service of a General Medical and Surgical Hospital. The ages of the patients ranged from 19 to 69 with a mean of 39.8.

The inventories were administered to the subjects in groups on four occasions, each 2 days apart, and counter balanced for order of presentation. The instructions to the subjects were "describe yourself."

## RESULTS

### *Comparability of Forms*

The question of comparability of the forms, aside from the issue of control of SD, is reflected in the intercorrelations of variables among the four forms. These intercorrelations represent the mean correlation between eight variables on two forms: for example, PA Form 1 with PA Form 2, etc. The mean intercorrelations ranged between .63 and .70; for $N = 39$, correlation of .40 is significantly different from 0 at the 1% level of confidence.

The level of the mean intercorrelations indicates a common pool of behavioral attributes are being tapped by the four forms. As a "comparable form" measure of test reliability, these mean intercorrelations appear to fall within the lower limits of such intercorrelations obtained generally from personality inventories. The homogeneity of the mean intercorrelations shows clearly that no one form departs importantly in pattern of tests results from the other forms.

TABLE 3

FREQUENCY OF SIGNIFICANT *t* TESTS BETWEEN FORMS BY VARIABLE MEANS

| Form | Form | | | Significance level |
|---|---|---|---|---|
| | 2 | 3 | 4 | |
| 1 | 2 | 1 | 2 | .05 |
| | 1 | 0 | 3 | .01 |
| 2 | | 1 | 1 | .05 |
| | | 0 | 3 | .01 |
| 3 | | | 2 | .05 |
| | | | 3 | .01 |

*Comparison of Control of SD in Four Forms*

The problem of the degree to which variable scores were confounded with or related to the SD factor was analyzed in three ways. First, each subject's SD score, as calculated from Form 3, was correlated with his variable scores on all four forms. The results indicate that no one variable on any of the forms is heavily confounded with SD. Only one of the 32 correlations involved reached a 1% level of significance. The mean and range of these correlations is comparable to those given in the manual of the Personal Preference Schedule (Edwards, 1953a). From these findings it is inferred that the methods of controlling SD used here are generally comparable and that they are as effective as the forced choice method used by Edwards.

While the correlational analyses so far reported indicate comparability of forms on proportional measures, mean variable scores could still be differentially affected by or related to SD. Analysis of variance and *t* tests were used to analyze the comparability of mean variable scores among the test forms. First, the means and distributions of variable scores were transformed to a common base, *T* scores. This was done by calculating the grand mean and standard deviation of the eight distributions of 39 scores each, obtained on each form. From these values, individual variable score distributions were transformed to *T* scores with a mean of 50 and standard deviation of 10. This was repeated on each form. Next, subjects were divided into High, Middle, and Low SD groups of equal *N*'s (13 in each group) according to their SD scores obtained on Form 3. Eight analyses of variance were carried out, one for each variable; three levels of SD, by four forms.

These analyses reveal few significant findings between forms, between SD groupings, or on the interaction of these two factors. Generally speaking, mean variable scores obtained from each of the test forms appear to be comparable estimates of the parameter from which they were drawn. Similarly, separating subjects according to their SD scores fails to reveal systematic differences with respect to mean variable scores across the four forms.

Despite the findings of the analyses of variance, the possibility remained that there may have been a tendency for one of the four forms to have a systematic rank position with regard to the other forms on mean variable scores. Analysis of the four together might obscure this trend. To explore this further, a more detailed analysis of mean variable scores was carried out by comparing each form with every other form by *t* test on each of the eight variables. The results of this analysis are summarized in Table 3.

It should be noted that if the four forms yield comparable mean variable scores, whatever significant mean differences occur, as measured by *t* test, should be approximately equally distributed among the four forms. The data given in Table 3 indicate this is not the case. A total of 48 tests were calculated, 24 involving each of the four forms. Table 3

TABLE 4

COEFFICIENT OF CONCORDANCE—*W*

| Forms | SD group | | |
|---|---|---|---|
| | High | Middle | Low |
| 1 | .03 | .30*** | .13 |
| 2 | .09 | .32*** | .10 |
| 3 | .10 | .28*** | .08 |
| 4 | .16* | .30*** | .19** |

* 5% level of significance.
** 1% level of significance.
*** .1% level of significance.

shows that a total of 19 $t$'s were significant, 9 at the 5% level, 10 at the 1% level. Of these, 14 involved Form 4 (5 at 5% level, 9 at 1% level), whereas only five were significant among the other three forms and four of these at only the 5% level. Both the number of significant $t$ tests and their distribution among the forms obviously departs from chance. Here then is clear evidence that so far as these conditions of measurement are concerned the four test forms studied show systematic differences. The one form, Form 4, yields mean variable scores differing significantly from the other three forms. These other forms show comparable means. Our data indicate that these differences between Form 4 and the other forms are most probably related to the SD factor. That is to say that Form 4 is the most susceptible to being influenced by variations in the balancing for the SD variable.

The comparability of Forms 1, 2, and 3 is again evident from the data given in Table 3. These data support the conclusion that in $Q$ sort-type measuring instruments, SD is controlled equally well when items of equal or varying SD values are permitted to compete for endorsement, so long as there is an overall balance between item pools for each variable.

The final method of analysis was to evaluate similarity of test profiles across the eight variables by the coefficient of concordance $W$ (Kendall, 1948). Comparisons were made for each of the High, Middle, and Low SD groups, form by form. These data are presented in Table 4.

The findings given in Table 4 show that all tests forms yield a significant amount of profile similarity for the Middle SD subjects but that only Form 4 shows homogeneity of profiles, though only slightly so, for the High and Low SD subjects. This indicates that in profile analysis too the check list form is slightly different from the $Q$ sort forms.

## DISCUSSION

The various methods of analysis reported here show consistently that the three $Q$ sort forms studied yield highly comparable data. As noted above, these findings indicate SD is controlled equally well when comparable or varying SD item values compete. Among the three forms of $Q$ sort used here, Forms 2 and 3 appear to have an advantage in that they permit calculation of an SD score independently of the variable scores. These SD scores can be used as an independent variable.

Although the data reveal some differences between the $Q$ sort and check list forms, these differences are slight. There is little data here to support the conclusion that the check list form produces clearly different test results in a practical sense. That differences, however slight, occur between Form 4 and the other three forms only in those variables less comparable with respect to MSD, supports the inference that SD item values affect variable scores and that Form 4, the check list, is more susceptible to this effect.

## SUMMARY

1. The potential utility of unconfounding rather than removing SD as a source of variance in personality inventories was developed.

2. Four forms replicating the same items drawn from an interpersonal check list were constructed to compare the effect of items of differing levels of SD competing for endorsement and to compare three variations of $Q$ sort construction with each other and with a check list.

3. The data support the conclusion that equal or varying item SD levels yield comparable results in variable scores and in control of SD.

4. The check list form was shown to differ slightly but consistently from the $Q$ sort forms in mean variable scores and in homogeneity of subject profiles. It was concluded, however, that these differences were so slight as to indicate the check list form has practical equivalence. The check list form was demonstrated to be more susceptible to minor variations in the balancing process.

## REFERENCES

BLOCK, J. A comparison of ipsative and normative ratings of personality. *J. abnorm. soc. Psychol.*, 1957, **54**, 50–54.

BROEN, W. E., JR., & WIRT, R. D. Varieties of response sets. *J. consult. Psychol.*, 1958, **22**, 237–240.

CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt.*, 1946, **6**, 475–494.

EDWARDS, A. L. *Manual for the Edwards Personal Preference Schedule.* New York: Psychological Corporation, 1953. (a)

EDWARDS, A. L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90–93. (b)

EDWARDS, A. L. Social desirability and probability of endorsement of items in the interpersonal check list. *J. abnorm. soc. Psychol.*, 1957, **55**, 394–396. (a)

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957. (b)

FORDYCE, W. E. Social desirability in the MMPI. *J. consult. Psychol.*, 1956, **20**, 171–175.

KENDALL, M. G. *Rank correlation methods.* London: Charles Griffin, 1948.

KOGAN, W. S., QUINN, R. D., AX, A. F., & RIPLEY, H. S. Some methodological problems in the quantification of clinical assessment by *Q*-array. *J. consult. Psychol.*, 1957, **21**, 57–62.

LA FORGE, R., & SUCZEK, R. F. The interpersonal dimension of personality: III. An interpersonal check list. *J. Pers.*, 1955, **24**, 95–112.

MEEHL, P. E., & HATHAWAY, S. R. The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.*, 1946, **30**, 525–564.

(Received October 31, 1960)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 31–37

# A COMPARISON AND REINTERPRETATION OF FACTOR STRUCTURES OF THE W–B, THE WAIS, AND THE HAWIE[1] ON AGED PERSONS

RUTH M. RIEGEL AND KLAUS F. RIEGEL[2]

*University of Michigan*

A number of factor analyses of the Wechsler Intelligence Scales (WAIS, W-B, and HAWIE) on adult or aged subjects have been conducted (Balinsky, 1941; Birren, 1952; Cohen, 1957; Goldfarb, 1941; Mundy-Castle, 1958; and Riegel, 1960). The results and their interpretations, however, have been contradictory. Birren identified four factors in the W-B responses of 60- to 74-year-old persons. These factors differed markedly from those reported by Cohen (1957) for almost the same age group (60 to 75+ years) on the WAIS. They corresponded closely, however, to Cohen's factor structures[3] of the age groups 18 to 19, 25 to 34, and 45 to 54 years. Cohen himself found that the factor structure for the oldest group differed noticeably from those for the three younger age groups and used this result as indication of changes in factor structure during later years of life. In light of Birren's findings this conclusion can be maintained only, if the discrepancy between W-B and WAIS factor structures of the age group 60 years and older *cannot* be clearly attributed to differences in the forms of the tests, and/or to particular differences in the techniques of rotation employed. Accordingly, the following problems suggested by the previous findings are to be reconsidered:

1. Does the factor structure of the Wechsler test, for the age group 60 years and older, vary among the different versions of the test?
2. Are there changes in the factor structure of the Wechsler test with advancing age, i.e., changes which may occur independently of any growth and decline in test performance?

Since recently the factor structure of the German version of the Wechsler test (HAWIE) has been made available our study is extended to a comparison which bears on the question:

3. Are there cultural differences between the factor structures of the Wechsler tests for the age group 60 years and older?

Factor analysis is a most useful tool where a large number of interrelated measures is to be reduced to a few underlying factors, a linear combination of which represents any of the measured variables. Since the Wechsler tests consist of 10 subtests only (or 11 when the vocabulary is included) factor analyses of this scale are of limited value. The most one can expect is information on the stability of the test structure between age groups, populations, and versions. One can hardly expect to learn much about the intellectual structure of older persons, since the number of subtests included in the battery is not large enough for such an analysis.

For this reason the purpose of our investigation is limited and represents mainly a critique and re-evaluation of studies in which such attempts have been made. A second purpose is the application of more appropriate methods for comparisons of different structures.

[1] *HA*mburg-*W*echsler-*I*ntelligenztest für *E*rwachsene.

[2] Section on Aging, Laboratory of Psychology, National Institute of Mental Health, Bethesda, Maryland. The preparation of this paper was completed by the senior author during the tenure of a Postdoctoral Fellowship from the Institute of Mental Health, United States Public Health Service. The authors are greatly indebted to J. E. Birren for his support and constructive criticism.

[3] The term factor structure refers to the configuration or pattern of factor loadings whereas changes in factor loadings that do not alter the interpretation of the factors will not be called structural changes.

## METHOD

### *Tests*

While the W-B and WAIS are sufficiently known, brief remarks on the HAWIE have to be made. The HAWIE is a translation of the W-B. Only some test items have been slightly modified. The standardization resembles that of Wechsler (1939) and was conducted by Anne Hardesty (Wechsler, 1956) on 1,831 persons of the age range from 10 to 60 years. The standardization of the HAWIE was extended and partially repeated for the age groups 55 to 75+ years by Riegel and Riegel (1959) and was based on the test results of 380 persons in Northern Germany. This extension resembles in many respects the standardization of the WAIS (Wechsler, 1955) for the older age groups by Doppelt and Wallace (1955). However, to most subjects only three or four alternate subtests were given, each combination allowing for an equally good estimate of the total test scores. The complete matrix of subtest intercorrelations was derived from a subgroup of 128 subjects to whom the whole test battery was administered and who closely resembled the total group of 380 subjects with respect to mean subtest performances. For a detailed discussion of the procedures applied, see Riegel and Riegel (1959).

### *Techniques of Comparisons*

Thurstone's (1947) method of rotation to simple structure is of particular value where psychological meaningful factor patterns are to be identified (Birren, 1952; Goldfarb, 1941; Mundy-Castle, 1958). However, when the main goal of the analysis is to compare different factor structures with each other, this technique is not satisfactory because no objective criteria are available to indicate when simple structure is attained. It is possible, therefore, that differences between factor matrices which have been rotated to simple structure are not inherent in the data but are merely due to differences in opinion as to when simple structure has been obtained. Among the methods available to yield comparable factor matrices are: Hotelling's principal components method (1933), and Ahmavaara's transformation analysis (1957).

The value of Hotelling's method lies in the fact that it yields unique solutions and thus enables the experimenter to compare the factor (component) structures of different groups. In an earlier investigation by one of us (Riegel, 1960) this method did not allow for a reliable identification of more than two factors (components), since these accounted for 65% of the total and for 90% of the common variance. In the studies under consideration more detailed factor matrices have been reported. In order to compare these matrices a method such as the transformation analysis of Ahmavaara has to be applied.

The method of Ahmavaara allows for a rotation of one matrix in such a way that it corresponds most closely to the matrix of comparison. The transformation matrix is obtained by the following formula:

$$L = (F'_{\mathrm{I}}F_{\mathrm{I}})^{-1}(F'_{\mathrm{I}}F_{\mathrm{II}})$$

Here, $L$ denotes the transformation matrix, which has to be normalized by rows before its application. $F_{\mathrm{I}}$ and $F_{\mathrm{II}}$ denote the factor matrices under consideration. Matrices rotated to simple structure (even to a slightly oblique one) may be compared by this method.

Ahmavaara demonstrated the correspondence between two transformed factor matrices by plotting all factor loadings of the two matrices against each other, and suggested the following index:

$$\Delta = \frac{\Sigma(F_{\mathrm{II}} - F_{\mathrm{I}}L)^2}{N}$$

This index, however, would falsely indicate a low congruence, if the factor loadings of one matrix are systematically lower than those of the other, but do not differ relative to each other. Because of this shortcoming Pearson's coefficient of correlation rather than Ahmavaara's index was used in our analysis to indicate the relationship between the different factor matrices.

## RESULTS

The three questions formulated above were tested by comparing a number of different factor structures reported in the recent literature. The factor matrix of the HAWIE was in each case the standard matrix of comparison. For each comparison a transformation matrix, $L$, was obtained. The matrices resulting from the multiplication of the factor matrix of the HAWIE with the particular transformation matrices, $L$, are supposed to correspond closest to the other factor matrices under consideration.

In particular, the HAWIE was compared with the factor structure of the WAIS (Cohen) and the W-B (Birren). In addition, an unrotated factor matrix of the WAIS for Cohen's age group of 70 to 74 years was derived by the authors. This seemed advisable since Cohen himself had used an oblique rather than orthogonal rotation. Finally, an unpublished factor analysis of the WAIS for the age groups 20 to 74 years was made available by Birren to the authors and thus the following structures were tested for similarity:

1. Factor structure of the HAWIE for the age group 65 to 75+ years (Riegel, 1960).

TABLE 1

FACTOR MATRIX OF THE HAWIE AFTER ORTHOGONAL ROTATION

| Subtests | A | B | C | D | E |
|---|---|---|---|---|---|
| General Information | .79 | .02 | .43 | .08 | .05 |
| Comprehension | .64 | .15 | .64 | −.04 | .06 |
| Digit Span | .55 | .41 | .23 | .08 | −.07 |
| Arithmetic | .54 | .10 | .43 | .32 | .16 |
| Similarities | .84 | .00 | .33 | −.06 | .20 |
| Coding | .75 | .34 | .03 | .18 | .05 |
| Picture Arrangement | .65 | .09 | .00 | .03 | .40 |
| Picture Completion | .75 | −.08 | .15 | .36 | .03 |
| Block Design | .75 | .21 | −.04 | .25 | .19 |
| Object Assembly | .50 | −.03 | −.08 | .47 | .27 |

Note.—All Subjects 65 to 75 years old.

Five centroid factors were extracted, and orthogonally rotated as shown in Table 1.

2. Factor structure of the W-B for the age group 60 to 74 years (Birren, 1952). Four factors were extracted and rotated to a slightly oblique simple structure. In Table 2 these factors are compared with the corresponding factors of the HAWIE, the latter being rotated to closest similarity with the first. The degree of similarity is revealed by a correlation coefficient of $r = .92$.

3. Factor structure of the WAIS for the age group 60 to 75 years (Cohen, 1957). This analysis is based on the data of Doppelt's (Doppelt & Wallace, 1955) extension of the standardization. The four factors identified after rotation to an oblique simple structure are compared in Table 3 with the corresponding factors of the HAWIE, the latter being rotated to closest similarity with the first. The degree of correlation is $r = .66$.

4. Factor structure of the WAIS for the age group 70 to 74 years (Riegel, present publication). The factor analysis is based on the correlation matrix reported by Doppelt and Wallace (1955). Four centroid factors were extracted by the present authors and are compared in Table 4 with the corresponding factors of the HAWIE, the latter being rotated to closest similarity with the first. The degree of correlation is $r = .93$.

5. Factor structure of the WAIS for the age groups 20 to 74 years (Birren, unpublished). This factor analysis is based on the standardization sample of the test (Doppelt & Wallace, 1955; Wechsler, 1955). Age and education were included as variables. The centroid factors were rotated to an orthogonal structure and are compared in Table 5 with the corresponding factors of the HAWIE, the latter being rotated to closest similarity with the first. The degree of correlation is $r = .96$.

All correlation coefficients summarized below are significantly different from zero ($p < .001$). The correlations do not differ significantly from each other ($p > .10$), except when compared with the correlation on the HAWIE and WAIS by Cohen ($p < .001$). These interpretations have to be accepted with some precautions, however, since the sampling distributions of such correlations

TABLE 2

COMPARISON OF THE TRANSFORMED HAWIE FACTOR MATRIX WITH THE W-B FACTOR MATRIX OF BIRREN HOLDING ROTATION CONSTANT

| | A | | B | | C | | D | |
|---|---|---|---|---|---|---|---|---|
| Subtests | HAWIE | W-B | HAWIE | W-B | HAWIE | W-B | HAWIE | W-B |
| General Information | .95 | .91 | .03 | −.02 | .01 | −.07 | .16 | .03 |
| Comprehension | .88 | .77 | −.24 | .10 | .06 | −.01 | .08 | .08 |
| Digit Span | .65 | .45 | .13 | .03 | .41 | .41 | −.01 | −.01 |
| Arithmetic | .80 | .68 | .07 | .01 | .30 | .32 | .34 | .19 |
| Similarities | .97 | .80 | .13 | .00 | −.10 | .05 | .22 | .33 |
| Coding | .76 | .66 | .46 | .51 | .41 | .31 | .17 | .00 |
| Picture Arrangement | .73 | .60 | .46 | .45 | .09 | .03 | .42 | .10 |
| Picture Completion | .78 | .72 | .35 | .37 | .13 | −.02 | .29 | .13 |
| Block Design | .76 | .55 | .57 | .38 | .35 | .38 | .34 | .29 |
| Object Assembly | .54 | .35 | .57 | .37 | .31 | .01 | .53 | .32 |

TABLE 3

COMPARISON OF THE TRANSFORMED HAWIE FACTOR MATRIX WITH THE WAIS FACTOR MATRIX OF COHEN HOLDING ROTATION CONSTANT

| Subtests | A | | B | | C | | D | |
|---|---|---|---|---|---|---|---|---|
| | HAWIE | WAIS | HAWIE | WAIS | HAWIE | WAIS | HAWIE | WAIS |
| General Information | .56 | .29 | .04 | −.09 | .11 | .41 | .43 | .10 |
| Comprehension | .57 | .39 | −.05 | .01 | .37 | .38 | .20 | −.10 |
| Digit Span | .14 | .00 | .19 | .09 | .32 | .45 | .26 | .03 |
| Arithmetic | .26 | .04 | .24 | .06 | .35 | .52 | .23 | .08 |
| Similarities | .73 | .42 | .11 | .10 | −.09 | .10 | .38 | −.02 |
| Coding | .21 | −.06 | .41 | .29 | .07 | .21 | .43 | .20 |
| Picture Arrangement | .51 | .30 | .44 | .20 | −.22 | −.09 | .22 | .10 |
| Picture Completion | .28 | .07 | .23 | .04 | −.05 | .01 | .58 | .38 |
| Block Design | .26 | .00 | .50 | .58 | −.07 | .09 | .43 | .00 |
| Object Assembly | .08 | −.01 | .53 | .56 | −.09 | −.03 | .34 | .03 |

among transformed factor matrices are unknown.

| | |
|---|---|
| HAWIE: W-B by Birren | .92 |
| HAWIE: WAIS by Cohen | .66 |
| HAWIE: WAIS by Riegel | .93 |
| HAWIE: WAIS by Birren | .96 |

## INTERPRETATION

The first question on differences in factor structures between different versions of the test, can be answered by comparing the correlation between HAWIE and W-B (Birren) with that between HAWIE and WAIS (Cohen). The latter is significantly lower than the first, however, it differs also significantly from the other correlations between HAWIE and WAIS (Birren, Riegel). Thus, it seems reasonable to conclude that the deviation of the factor structure of the WAIS (Cohen) from the other factor structures, is due to differences in opinion as to when simple structure is obtained and to the rotation procedures applied rather than to differences in the different versions of the test and inherent in the data.

Since no comparisons with the factor structures of younger age groups have been included in the present analysis the second question on changes in factor structure with advancing age, as recently proposed by Cohen (1957) and earlier by Balinsky (1941), Garrett (1946), Spearman (1927) and others, can be answered tentatively only. Relying on our first conclusion, it is again

TABLE 4

COMPARISON OF THE TRANSFORMED HAWIE FACTOR MATRIX WITH THE WAIS FACTOR MATRIX OF RIEGEL HOLDING ROTATION CONSTANT

| Subtests | A | | B | | C | | D | |
|---|---|---|---|---|---|---|---|---|
| | HAWIE | WAIS | HAWIE | WAIS | HAWIE | WAIS | HAWIE | WAIS |
| General Information | .75 | .78 | .33 | .34 | .03 | −.16 | .04 | .14 |
| Comprehension | .70 | .67 | .34 | .20 | −.25 | −.26 | −.08 | −.16 |
| Digit Span | .60 | .66 | −.06 | .06 | −.20 | −.35 | .10 | .08 |
| Arithmetic | .69 | .75 | −.02 | .09 | −.28 | −.26 | .00 | .06 |
| Similarities | .74 | .75 | .34 | .32 | .25 | .16 | −.12 | −.24 |
| Coding | .72 | .77 | −.15 | −.22 | .10 | .15 | .10 | .27 |
| Picture Arrangement | .59 | .66 | −.08 | .11 | .34 | .40 | −.23 | −.08 |
| Picture Completion | .68 | .77 | .11 | .23 | .15 | .23 | −.08 | .28 |
| Block Design | .72 | .63 | −.20 | −.38 | .22 | .27 | .04 | −.18 |
| Object Assembly | .55 | .57 | −.20 | −.52 | .14 | −.12 | .05 | .06 |

TABLE 5

COMPARISON OF THE TRANSFORMED HAWIE FACTOR MATRIX WITH THE WAIS FACTOR MATRIX OF BIRREN HOLDING ROTATION CONSTANT

| Subtests | A | | B | | C | | D | |
|---|---|---|---|---|---|---|---|---|
| | HAWIE | WAIS | HAWIE | WAIS | HAWIE | WAIS | HAWIE | WAIS |
| General Information | .86 | .81 | .09 | .27 | .38 | .26 | −.11 | −.03 |
| Comprehension | .87 | .71 | .27 | .19 | .20 | .26 | −.01 | −.17 |
| Digit Span | .63 | .59 | −.15 | −.05 | .08 | .14 | .23 | .29 |
| Arithmetic | .80 | .66 | .09 | .17 | .51 | .26 | .21 | .27 |
| Similarities | .93 | .72 | −.04 | .04 | .42 | .24 | −.24 | −.13 |
| Coding | .77 | .68 | −.40 | −.42 | .34 | .07 | .16 | .04 |
| Picture Arrangement | .80 | .62 | −.37 | −.28 | .54 | .31 | −.09 | −.08 |
| Picture Completion | .72 | .60 | −.12 | −.14 | .59 | .49 | −.03 | −.09 |
| Block Design | .80 | .51 | −.45 | −.26 | .54 | .57 | .11 | .08 |
| Object Assembly | .60 | .44 | −.33 | −.28 | .73 | .49 | .15 | .01 |

quite possible that the differences in factor structures between age groups as reported by Cohen (1957) are due to the procedures used and not inherent in the data. This is especially likely since there is good correspondence in factor structure between Cohen's younger age groups and those reported for old subjects by Birren (1952) and Riegel (1960). Thus, only Cohen's old age group deviates in factor structure from his younger groups. These comparisons have not been made, however, by applying Ahmavaara's technique.

Further evidence to the second question can be derived from Birren's unpublished study of the WAIS in which subjects ranged in age from 20 to 74 years. The factor structure of this larger age group deviates as little from that of the HAWIE for the age group 65 to 75+ years ($r = .96$), as the factor structure of the WAIS (Riegel) for the age group 70 to 74 years ($r = .93$). Finally, also the factor structures of the HAWIE as determined by using Hotelling's principal component method did not reveal structural differences between the age groups 20 to 34, 35 to 49, 50 to 64, and above 65 years (Riegel, 1960).

In answering the third question on differences in factor structure between different populations (American and German) we have to rely again on the above correlations. In view of the high similarity between HAWIE and W-B (Birren) as well as between HAWIE and WAIS (Birren and Riegel) it seems safe to conclude that no differences are revealed by administering the Wechsler test to different populations.

Summarizing these findings, our conclusions can be considerably simplified. All one can conclude is that there are no differences in factor structures between different versions of the Wechsler test, between different populations and seemingly between different age groups. The only deviation in factor structure of the WAIS as reported by Cohen seems to be due to the rotation procedures applied and possibly to differences in opinion as to when simple structure is obtained.

## DISCUSSION

The results suggest that the methods of extraction and rotation influence the factor structures. In our opinion, the criteria for simple structure are not sufficient to determine factor structures that are to be compared with each other. Differences noted may be primarily due to differences in opinion as to when simple structure is obtained, rather than to differences inherent in the data. In particular, this will be true if factors are rotated to an oblique rather than orthogonal structure, since here additional subjective clues are influencing the decisions that have to be made. Generally, greater precaution ought to be observed in the interpretation of factors but especially of differences between factor structures. The transformation analysis seems

to be a useful analytic method to warrant comparability and thus represents a means of integrating the multitude of factor analytical findings. The methods of rotation by electronic computers may also serve this purpose. These methods yield unique solutions but only approximations of simple structure.

It is unlikely that differences in factorial structures of the Wechsler tests between age groups can be detected. Only few factors were identified which represent the most general characteristics of the test such as general intellectual, verbal-performance, numercial, or speed aspects. As revealed by an earlier analysis (Riegel, 1960) the most one can detect in such comparisons are differences in factor loadings of subtests and thus cues about changes in their diagnostic value. Since it has to be also realized that each subtest is composed of a number of different variables, more distinct results for a psychology of aging may be obtained by analyzing the subtests themselves taking the item intercorrelations into account.

Generally, the failure to find differences in factor structures between different groups and conditions may be due to restrictions in the range of abilities sampled by the Wechsler tests. Accordingly, one common factor accounts for the major part of the variation and little variability is left to more specific factors that could be expected to differ between groups and conditions. Our principal component analysis (Riegel, 1960) yields clear evidence to this point.

No cultural differences in the factor structures of the Wechsler tests were noticed. Thus, in spite of some changes in items the main features seem to have been maintained in the German version of the test. However, more detailed investigation of cultural differences should be based again on item statistics because some subtests of the W-B, the WAIS, and the HAWIE differ in length from each other. Crude comparisons of the rate of growth and decline have been made between the two populations by using standard scores, i.e., expressing the mean subtest scores within one age group in terms of the highest average achievements for each subtest, namely the achievements of the age group 20 to 34 years. These comparisons on the W-B and HAWIE (Riegel, 1958) and on the WAIS and HAWIE (Birren, 1959) have not shown any marked differences between the American and German populations for the age range from 20 to 75+ years.

## SUMMARY

Factor analyses conducted on the WAIS and W-B (age groups 60 to 75+) have indicated: (*a*) WAIS factor structure (Cohen) differs from W-B factor structure (Birren). (*b*) WAIS factor structure of the age group 60 to 75+ differs from those for the younger age groups (Cohen). In the present study these findings were reconsidered, applying more suitable methods for comparisons (Ahmavaara). (*c*) W-B and WAIS were compared with HAWIE factor structure in order to analyze cultural differences.

The results failed to show: (*a*) Differences in factor structure between different versions of the test for the age groups above 60. (*b*) Sufficient evidence for changes in factor structures with advancing age. (*c*) Differences in factor structures of the American and German populations.

## REFERENCES

AHMAVAARA, Y. On the unified factor theory of mind. Helsinki: *Ann. Akad. Sci. Fennicae, Ser. B.*, 1957, **106**.

BALINSKY, B. An analysis of the mental factors in various age groups from nine to sixty. *Genet. psychol. Monogr.* 1941, **23**, 191-234.

BIRREN, J. E. A factorial analysis of the Wechsler-Bellevue Scale given to an elderly population. *J. consult. Psychol.*, 1952, **16**, 399-405.

BIRREN, J. E. Principles of research on aging. In J. E. Birren (Ed.), *Handbook of aging and the individual.* Chicago: Univer. Chicago Press, 1959.

COHEN, J. The factorial structure of the WAIS between early adulthood and old age. *J. consult. Psychol.*, 1957, **21**, 283–290.

DOPPELT, J. E., & WALLACE, W. L. Standardization of the Wechsler Adult Intelligence Scale for older persons. *J. abnorm. soc. Psychol.*, 1955, **51**, 312–330.

GARRETT, H. E. A developmental theory of intelligence. *Amer. Psychologist*, 1946, **1**, 372–378.

GOLDFARB, W. An investigation of reaction time in older adults. *Teach. Coll. Contr. Educ.*, 1941, No. 381.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. *J. educ. Psychol.*, 1933, **24**, 417–441.

MUNDY-CASTLE, A. C. Electrophysiological correlates of intelligence. *J. Pers.*, 1958, **26**, 184–199.

RIEGEL, K. F. Ergebnisse und Probleme der psychologischen Alternsforschung. Teil 1. *Vita Hum.*, 1958, **1**, 52–64, 111–128.

RIEGEL, RUTH M. Faktorenanal,se des Hamburg-Wechsler-Intelligenztests für Erwachsene (HAWIE) für die Altersstufen 20–34, 35–49, 50–64 und 65 Jahre und älter. *Diagnostica*, 1960, **6**, 41–66.

RIEGEL, RUTH M. & RIEGEL, K. F. Standardisierung des Hamburg-Wechsler-Intelligenztests für Erwachsene (HAWIE) für die Altersstufen über 50 Jahre. *Diagnostica*, 1959, **5**, 97–128.

SPEARMAN, C. *Ability of man*. London: Macmillan, 1927.

THURSTONE, L. L. *Multiple factor analysis*. Chicago: Univer. Chicago Press, 1947.

WECHSLER, D. *The measurement of adult intelligence*. Baltimore: Williams & Wilkins, 1939.

WECHSLER, D. *Manual for the Wechsler Adult Intelligence Scale*. New York: Psychological Corporation, 1955.

WECHSLER, D. *Die Messung der Intelligenz Erwachsener*. Bern: Huber, 1956.

(Received November 9, 1960)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 38–43

# SIMILARITY AND SUCCESS IN THERAPEUTIC DYADS[1]

ROBERT C. CARSON AND RALPH W. HEINE
*Duke University* *University of Chicago*

Despite repeated reference in the literature of psychotherapy to the importance of the nature of the patient-therapist relationship in determining the course and outcome of psychotherapy, extremely few studies on the problem have been reported. Many investigators have sought to relate various personality variables of patients to success in therapy; some, but many fewer, similar studies have been made of therapists. Studies of patient-therapist "fit" as it affects outcome, however, have rarely been undertaken. The present study is of this dyadic type, in that it is concerned with the effect of patient-therapist personality similarity upon therapeutic success.

The relationship between general patient-therapist personality similarity and therapeutic success has been studied previously by Axelrod (1951), Tuma and Gustad (1957), and Schopler (1958) using various definitions of "similarity" and "success." While the results of none of these studies are unequivocal, only Schopler failed to find evidence for a positive relationship. Heine and Trosman (1960) and Orlinsky (1959) report a positive correlation between similarity of patient and therapist expectations of the relationship and duration of therapy. Vogel (1959) obtained partial confirmation of his hypothesis that similarity with respect to authoritarian behavior and attitudes of patient-therapist pairs would enhance their degree of satisfaction with the relationship. In several unpublished studies, however, the present authors, working on the seemingly well-founded assumption that global patient-therapist personality similarity and therapeutic success would be related in linear fashion, obtained generally disappointing results. A subsequent rethinking of the problem led to a further analysis of the data from a different standpoint.

We reasoned that if, as has often been suggested, psychotherapeutic success depends upon the therapist's being able to achieve an optimum balance between empathy and objectivity in dealing with his patient, then the relationship between patient-therapist similarity and therapeutic success might well be of curvilinear form. With very high similarity the therapist might be unable to maintain suitable distance and objectivity, whereas in the case of great dissimilarity he would not be able to empathize with or understand the patient's problems; in either case one might expect a decrement in therapeutic success. Formally stated, our hypothesis in the present study was as follows: As patient-therapist personality dissimilarity increases through a range from maximum to minimum congruence, there will be a corresponding increase in therapeutic success up to a certain (unspecified) point, beyond which further increases in dissimilarity will be associated with decreasing success of the therapeutic relationship.[2]

## METHOD

Senior medical students at the University of Chicago School of Medicine are required as part of their training in psychiatry to take an 18-week course which consists primarily of treating psychotherapeutically, under supervision, a selected patient who has been referred to the Psychiatry Clinic. Our subjects were patients and therapists drawn from this program during four successive academic quar-

[1] This study was supported in part by a grant from the Psychiatric Training and Research Fund of the Illinois Department of Public Welfare. The authors wish to express their appreciation to John Altrocchi and Martin Lakin for their critical comments and editorial assistance.

[2] The curvilinear form of the relationship was not obvious in the few scatter plots originally constructed from more limited data. The specific hypothesis tested was not a direct product of these original data analyses, but was temporally prior to the analysis upon which this paper is based. Schopler (1958), whose work came to our attention after the completion of the present study, had proposed this hypothesis originally in his study. However, his findings led him subsequently to hypothesize a relationship of a rather different form, based upon certain tenuous trends in his data.

ters. From a total population of some 70 patient-therapist pairs, we were able to collect all data relevant to this study on 60 cases. All patients were told that their treatment would be limited to 18 weeks, at the end of which period they would either be terminated or referred elsewhere. With rare exceptions, patients were seen once a week for their psychotherapy interviews.

The patient sample contained 35 females and 25 males, with an average age of about 38 years. The range of disorders for which the patients had sought treatment is probably not atypical for an adult psychiatric out-patient clinic, with the exception that psychotic disorders probably were underrepresented by virtue of the program's selection process. With very rare exception neither patient nor therapist had had any prior direct experience with formal psychotherapy. The therapists, of whom eight were female, were unselected except for their status as fourth-year medical students.

The MMPI was used to assess similarity of personality between patient and therapist. All patients and therapists were given the booklet form of the MMPI at the beginning of treatment. Using fairly liberal validity scale criteria ($L \leq 70$, $F \leq 80$, $K \leq 75$), no profile in either group was considered invalid. Although there are many possible methods of deriving global indices of similarity between two MMPI profiles (Cronbach & Gleser, 1953), Hathaway and Meehl, after extensive investigations of the relative validity of various methods, concluded that simple rank-order correlation of the two sets of *T* scores is probably the most efficient index available.[3] This was the method of similarity assessment chosen for the present study. It is important to note that this method will yield a measure of similarity of profile *shape* exclusively, since the variance associated with elevation and scatter is discounted in the computational process. In terms of Cronbach's (1958) component analysis of dyadic similarity measures, profile means and standard deviations are automatically equalized, and only the residual correlational component determines the magnitude of the similarity index.

The specific operations used in computing the similarity index were as follows. All profiles were *K*-corrected in the usual manner at the outset. The *Mf T* score for each female subject (whether a patient or one of the eight female therapists) was transformed into a *T* score representing the same amount of deviation from the standardization mean as had been originally obtained, but in the opposite direction. Thus a high *Mf T* score for a female subject, instead of representing a deviation in the direction of excessive "masculinity," now represented a deviation in the direction of excessive "femininity." This procedure had the effect of stabilizing the meaning of the *Mf* score for both males and females. The ? validity score was discarded as superfluous because of the virtual absence of variability in this score in our sample of profiles. *T* scores for all other scales appearing on the standard profile form, including *L*, *F* and *K*, were ranked for all patients and therapists. For each patient-therapist pair, differences between ranks for each of the 13 scales were obtained by subtraction and squared. The index of similarity was the sum of these squared rank differences for each pair.[4] These indices, which, of course, are inversely proportional to degree of profile similarity, ranged from 31.5 to 614.5 in the 60 pairs.

The criterion for assessing success of therapy involved a composite of several outcome ratings made at the end of the period of treatment by the supervising psychiatrist on each case. A total of eight senior staff members served as supervisors in the program and were assigned to patient-therapist pairs on an unsystematic basis. The supervisors were asked to make the ratings as part of a questionnaire concerned with a variety of variables under investigation in the larger project of which this study was a part. So far as is known, none had any knowledge of the MMPI profiles of either the patients or the therapists whose cases they supervised, nor were any aware that their ratings would be used to test the hypothesis which is the subject of this paper.

Ratings on five items comprise the composite criterion. Two of the items asked the supervisor to rate (*a*) "degree of satisfaction of the patient with outcome of therapy," and (*b*) "your rating of the outcome of therapy" on nine-point scales. The remaining three items were of a more complex nature and asked for a rating of the patient's status with respect to three important dimensions at both the beginning and end of therapy; these dimensions were (*a*) occupational adjustment, (*b*) adequacy of interpersonal relations, and (*c*) symptomatic status. These items were arranged on a five-point scale from poor functioning to good, each of the intermediate as well as the end points having an appropriate descriptive word or phrase to delimit their meaning precisely. The score on each of these more complex items was taken as the "end" rating minus that of the "beginning," thus providing a measure of amount of positive change. Pearson *r*'s for the intercorrelations among all five items ranged from .32 to .73, the average value being .46. The overall "success" index was simply the algebraic sum of the scores on the above five items; it ranged from 2 to 20 points, with a mean of 11.63 for the 60 cases.

The therapists in the project were asked to complete the same research questionnaire as was given to the supervisors. While it did not seem appropriate in the present study to use these therapist ratings as criteria of success, they do provide some basis for evaluating the reliability of the supervisor's ratings. Forty-eight of the 60 therapists

[3] Hathaway, S. R. Personal communication. July 9, 1959. It may be noted that the emphasis given to profile shape by this index is in keeping with a trend toward emphasis on shape in the clinical interpretation of MMPI profiles.

[4] Since *N* is constant, computation of rho would be superfluous.

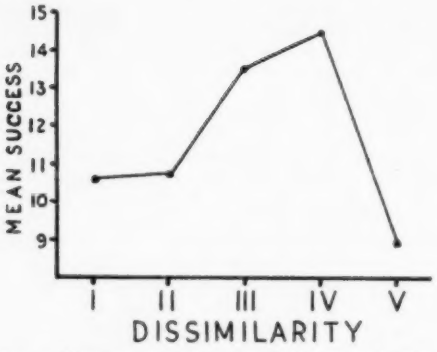


FIG. 1. Mean composite success rating as a function of patient-therapist similarity index.

provided ratings on all of the component items of the success index. The interrater reliability coefficient of the global success score, based upon these 48 cases, was .55; this is highly significant statistically, although it obviously accounts for only a portion of the total variance of the ratings.

## RESULTS

The 60 pairs were first categorized according to score on the MMPI similarity index. As a compromise to achieve a reasonable number of categories, each comprising a reasonable number of pairs, it was decided to establish five groups of 12 pairs each. The 12 pairs with the lowest indices ($< 265$) were arbitrarily designated Group I, the "very high" similarity group. The next 12 pairs in order of increasing scores on the index (265–318) became the "moderately high" similarity group and were designated Group II. Group III, comprising the next 12 pairs (indices = 319–386), was the "medium" similarity group. Group IV ("moderately low" similarity; indices = 387–456) and Group V ("very low" similarity; indices = $> 456$) were established in like manner.

The mean therapeutic success score, based on supervisor's ratings as described above, was computed for the pairs in each of the five groups; these means are presented graphically in Figure 1. It can be seen from Figure 1 that the form of the relationship between patient-therapist personality similarity and therapeutic success is curvilinear, thus supporting the hypothesis. With increasing dissimilarity there is at first an increase in therapeutic success, but with further increases in dissimilarity beyond an optimal point there is a precipitous decline in the rated success of the therapeutic relationship.[5] A simple analysis of variance was performed on these data, resulting in an $F$ of 3.06, which is significant beyond the .05 level for 4 and 55 degrees of freedom.[6] Supplementary $t$ tests revealed that Group IV, the group having the highest mean success rating, was significantly superior in rated success to either Group I ($t = 2.38$; $p < .05$) or Group V ($t = 3.90$; $p < .01$) at the extremes of the similarity dimension.

The interpretation of this finding, based as it is upon a global index of profile similarity, requires a cautious approach along the lines suggested by Cronbach's (1958) carefully reasoned analysis of such measures. Particularly relevant in this context is the possibility that the obtained relationship could be accounted for wholly on the basis of profile variations in one or another of the two classes of subjects, patients or therapists. Under such a circumstance a dyadic interpretation would be nonparsimonious. Suppose, for example, that the MMPI profiles of therapists were to vary but little in shape from a "constant" pattern reflecting a defensive set, while the profiles of patients might show marked variability in terms of similarity to this "defensive" pattern; the profile similarity index would then be essentially a measure of patient defensiveness, which might itself vary with success in a curvilinear fashion. Note, however, that this reasoning requires that there will be systematic variations in the shapes of profiles for patients at the various levels of patient-therapist similarity, and unless this con-

[5] It might be noted that none of the five groups of patient-therapist pairs were markedly unsuccessful as judged from average success scores. Possibly this results from a systematic bias of the raters to upgrade the success of the relationships they supervised. On the other hand, however, we have in fact been surprised by the unexpectedly high level of overall success obtained by patient-therapist pairs in this program.

[6] This is equivalent to a correlation ratio (eta) of .43 for the regression of success scores on similarity.

TABLE 1

MEAN MMPI *T* SCORES FOR PATIENTS AND THERAPISTS IN EACH EXPERIMENTAL GROUP

| | Groups | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | I | | II | | III | | IV | | V | |
| Scales | Patient | Therapist | Patient | Therapist | Patient | Therapist | Patient | Therapist | Patient | Therapist |
| *L* | 50.17 | 44.75 | 47.92 | 47.00 | 50.25 | 47.00 | 47.75 | 46.08 | 50.25 | 43.83 |
| *F* | 55.42 | 52.50 | 61.42 | 51.33 | 57.42 | 49.00 | 60.75 | 52.58 | 61.92 | 51.33 |
| *K* | 60.25 | 60.08 | 49.42 | 60.50 | 55.83 | 62.75 | 53.08 | 61.33 | 51.33 | 61.33 |
| *Hs* | 60.58 | 50.42 | 64.58 | 52.42 | 66.17 | 49.42 | 67.25 | 50.25 | 72.08 | 49.08 |
| *D* | 68.42 | 58.00 | 73.17 | 56.50 | 78.00 | 52.50 | 74.25 | 50.83 | 78.33 | 51.92 |
| *Hy* | 66.42 | 58.33 | 66.08 | 57.17 | 74.92 | 56.92 | 71.50 | 55.83 | 71.67 | 54.92 |
| *Pd* | 68.50 | 60.00 | 69.67 | 60.00 | 75.25 | 59.67 | 67.17 | 61.17 | 63.92 | 59.67 |
| *Mf* | 56.00 | 61.42 | 53.25 | 61.83 | 55.92 | 62.92 | 48.17 | 65.25 | 52.25 | 62.58 |
| *Pa* | 56.17 | 56.17 | 59.75 | 53.67 | 70.00 | 51.75 | 58.17 | 56.25 | 64.25 | 54.75 |
| *Pt* | 64.17 | 59.08 | 68.75 | 55.67 | 74.50 | 53.75 | 68.33 | 50.67 | 74.67 | 52.92 |
| *Sc* | 60.67 | 56.33 | 69.25 | 53.42 | 69.83 | 53.67 | 67.08 | 55.92 | 70.00 | 54.25 |
| *Ma* | 49.58 | 57.08 | 57.08 | 55.58 | 55.92 | 56.92 | 54.08 | 60.08 | 52.08 | 60.17 |
| *Si* | 51.42 | 48.58 | 60.42 | 49.75 | 60.17 | 49.17 | 58.42 | 46.08 | 60.50 | 48.58 |

dition were met one would be obliged to reject the proposed monadic interpretation of the results. In order to evaluate this kind of interpretation, as well as the monadic interpretation corresponding to the less probable case of therapist's profiles varying in similarity to a "common" patient profile, the "average" profile for both patients and therapists in each of the five groups was constructed from mean scores on each scale. These mean scores are presented in Table 1. The scales on each of these average profiles were subsequently ranked in the same manner used to compute the similarity indices. Profile shape, as represented in these rankings, appears to be strikingly similar across the five groups for patients as well as therapists. This impression was confirmed by computing a coefficient of concordance ($W$) for each of the two sets of ranked profiles; $W$ was .88 for the patients and .91 for the therapists (Maximum value = 1.00). We may conclude, then, that neither the patients nor the therapists *in general* show systematic differences in profile shape in the five groups.

It might be observed that there are unexpectedly substantial variations in elevation among the five mean profiles of patients. However, these variations do not appear to be related in any regular fashion to degree of patient-therapist profile similarity, nor is it easy to conceive of how they could account, on a monadic basis, for the results obtained. Relatively, patients in Group I, who have low profile elevation, had an unsuccessful therapy experience, while patients in Group IV, who also have low elevation, had a successful experience. On the other hand, patients in Group III, who have high elevation, had successful relationships, while patients in Group V, also having high elevation, had unsuccessful relationships. Thus the data will not readily support a monadic interpretation of the results, and the view is justified that we are dealing with a truly dyadic phenomenon.

## DISCUSSION

Perhaps the most obvious question concerning the results of this investigation is the extent to which they can be generalized. The patient-therapist relationships studied here were atypical in two respects: (*a*) the marked inexperience of the therapists, and (*b*) the time-limited (and necessarily goal-limited) orientation of the treatment. The question of generality is clearly an empirical one, calling for replication of the study in a more typical psychotherapeutic setting. It might be noted, however, that our findings parallel those of Fiedler (1958), who concludes from his work with task groups that group effectiveness is related in curvilinear fashion to the "psychological distance" of the group leader. Insofar as the task group and the therapeutic dyad are comparable there is a basis for assuming some generality of the phenomenon. Moreover, the fact that the therapists in this study were essentially untrained suggests that the phenomenon is somehow basic to personal interaction in the therapeutic context, and that whatever are the effects of specific psychotherapeutic training they must operate with the personality similarity factor as a baseline. While it seems unlikely that the acquisition of psychotherapeutic knowledge and skills could render the similarity variable totally irrelevant, the extent to which training would alter the pattern obtained of course remains a fundamental question. For the present, in any case, it can be stated as a working hypothesis that the effect demonstrated represents something stable in patient-therapist relationships. Confirmation of this hypothesis could make possible the development of logical, objective, actuarially sound procedures for pairing patients with therapists, an exciting idea.

The results give strong support to the views of interpersonally oriented psychotherapists who emphasize the importance of the nature of the interaction between patient and therapist as persons in determining the process and outcome of the therapeutic relationship. Fromm-Reichmann (1949), for example, has been very explicit on this point, suggesting that therapists should refrain from working with individuals whose personalities are markedly out of tune with their own. The problems have come in attempting to specify in advance with precision who should and who should not get together in a psychotherapeutic relationship. The identification of a measurable interactive dimension which

may be related to therapeutic outcome in a specifiable way (i.e., global personality similarity) could be a step toward achieving what is needed in this regard.

Turning to the question of what these results mean in terms of the therapeutic process, it is perhaps not surprising, in view of the rather gross nature of our approach, that not a great deal has been really clarified. The reasoning underlying our hypothesis was admittedly vague and inelegant—essentially we assumed that excessive similarity might make the therapist ineffective by virtue of overidentification with the patient and his problems, and that excessive dissimilarity might have the same result by making it impossible for him to understand or appreciate what the patient was trying to say to him. Fiedler (1958) explains his findings on task group leadership in somewhat similar terms. While the range of possible explanatory constructs for the phenomenon demonstrated is very broad, we suspect that in essence most of them will reduce to something very close to this rather simple formulation. At present there appears to be little better to offer.

The conceptual difficulties encountered in attempting to come to grips with the result of this study stem in part from its inherent limitations, in particular our enforced reliance on a global index of similarity, and in part from the not unrelated problem of having very little knowledge upon which to base our thinking about what are the important dimensions of the therapeutic relationship. We hope and expect that further research will resolve these issues.

## SUMMARY

An index of personality similarity for patient-therapist pairs was derived from scores made by a group of 60 patients and their therapists on the MMPI, and an attempt was made to relate similarity to success in therapy, as measured by supervisor's ratings. Success was found to vary significantly with similarity, the form of the relationship being curvilinear, in accord with the investigators' hypothesis that either extreme similarity or extreme dissimilarity would impede the therapeutic process. The results were discussed in terms of their generality, and their practical and theoretical implications.

## REFERENCES

AXELROD, J. An evaluation of the effect on progress in therapy of similarities and differences between the personalities of patients and their therapists. Unpublished doctoral dissertation, New York University, 1951.

CRONBACH, L. J. Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univer. Press, 1958. Pp. 353–379.

CRONBACH, L. J., & GLESER, G. C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50**, 456–473.

FIEDLER, F. E. Interpersonal perception and group effectiveness. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univer. Press, 1958. Pp. 243–257.

FROMM-REICHMANN, FREIDA. Notes on personal and professional requirements of a psychotherapist. *Psychiatry*, 1949, **12**, 361–378.

HEINE, R. W., & TROSMAN, H. Initial expectations of the doctor-patient interaction as a factor in continuance in psychotherapy. *Psychiatry*, 1960, **23**, 275–278.

ORLINSKY, NANCY. Patients' and therapists' definitions of therapist's role in relation to outcome of therapy. Unpublished doctoral dissertation, University of Chicago, 1959.

SCHOPLER, J. H. The relation of patient-therapist personality similarity to the outcome of psychotherapy. Unpublished doctoral dissertation, University of Colorado, 1958.

TUMA, A. H., & GUSTAD, J. W. The effects of client and counselor personality characteristics on client learning in counseling. *J. counsel. Psychol.*, 1957, **4**, 136–141.

VOGEL, J. L. Authoritarianism in the therapeutic relationship. Unpublished doctoral dissertation, University of Chicago, 1959.

(Received November 18, 1960)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 44–49

# SEXUAL SYMBOLS, STRUCTURED AND UNSTRUCTURED

KEN LESSLER [1]

*Michigan State University*

According to Freudian theory (1938) there is a symbolic association between the male and female genitalia and certain classes of objects which are similar to them in shape. The experimental investigation of the Freudian hypothesis was initiated by Levy (1954), and has been extended by Starer (1955), Jones (1956), Winter and Prescott (1957), and Moos and Mussen (1959). Although Levy, using children as subjects, did not find support for the Freudian hypothesis, positive results were reported by later investigators using adult subjects. Jones found that males identified more symbols in a manner consistent with the Freudian hypothesis than did females. In contrast to the findings of Levy and Jones, most psychoanalytic writers (e.g., Erikson, 1950; Fenichel, 1945) believe that children of *both* sexes are aware of biological sex differences as early as three or four years of age. To this writer's knowledge, neither Jones's nor Levy's results have been replicated.

The study by Levy, and those which followed, investigated the association suggested by Freud between the shape of the symbol and its purported referent, the sexual organ. It is suggested by this author that a relatively shapeless stimulus, such as a texture, may also be adequate to symbolize some aspects of masculinity and femininity. Rabin (1959) found that the "mother" and "father" cards on the Rorschach were perceived differently in terms of their textures, i.e., Card IV was perceived as heavy and rough, whereas Card VII was perceived as light and smooth.

The present writer sought to explore both the symbolic significance of structured (shaped) symbols, with special reference to age and sex differences, and to introduce an unstructured stimulus, a texture, as another possible mode of symbolic expression of masculinity and femininity. Specifically the study was conducted to answer the following questions:

1. Are structured symbols sorted as predicted by the Freudian hypothesis? Is the response to the structured symbols affected by the age and sex of the respondent?

2. Can textures be used to symbolize masculinity and femininity, i.e., are the dark and rough textures more often identified as masculine, and are the light and soft textures more often identified as feminine than would be expected by chance? Are there age and sex differences in the responses to the texture symbols?

## METHOD

*Sample.* The sample was composed of 120 subjects: 20 male, and 20 female subjects at the fourth-, ninth-, and college grade levels with mean ages of 9–4, 14–4, and 20–2, respectively. Although the subjects were not individually matched for intelligence and socioeconomic background, the three groups were believed to be similar in these respects. The grade school and high school populations from which the subjects were drawn were highly saturated with the children of university faculty and students.

*Materials.* Twenty structured and 20 unstructured (texture) symbols were used. There were 10 male and 10 female symbols of each type.[2] The structured symbols were drawn on $3 \times 5$ cards and the texture symbols on $3 \times 3$ cards. The symbols were selected from a larger group which was submitted to five judges well versed in psychoanalytic theory. A symbol was included in the final set if at least four out of five judges agreed upon its symbolic significance, and if it added to the overall variety of the symbol set. The structured symbols presented to the judges included those used by Levy (1954) and Starer (1955), as well as other symbols created by the writer so that the figures could not be sorted on the basis of sharpness vs. roundness. The final group of structured symbols included one symbol used by Levy and five of those used by Starer. The texture symbols presented to the judges were created by the writer and varied generally on the dark-light and rough-smooth continua.

[1] The author wishes to express his appreciation to T. Allen, G. Anderson, H. H. Anderson, D. G. Brown, A. I. Rabin, and S. Siegel for their helpful criticisms of the study, and to the public schools of East Lansing, Michigan for their cooperation in making subjects and facilities available.

[2] Copies of the symbols are available from the author upon request.

*Procedure.* Each subject was seen individually by the author and was instructed to sort cards into one of two piles as follows:

> I have some cards with figures or pictures on them. I would like you to place each card in the pile to which you think it belongs. If it reminds you more of a father, brother, boy, or man, place it in this (pointing) pile. If it reminds you more of a mother, woman, girl, or sister, place it in this pile. They will not look like anything, so you will have to guess. Please turn the cards over when you put them down.

Two labels were placed in front of the subject. On one label was printed "father, brother, boy, man," and on the other, "mother, sister, girl, woman." A small achromatic picture of a man and boy and one of a girl and woman were placed above the labels at the head of the corresponding piles for the fourth-grade subjects. The order of presentation of the structured and texture cards, and the side on which the male and female labels were placed, were randomized. The pile in which the cards were sorted was recorded by the experimenter while the subject was sorting. The score for each subject was the number of cards sorted in the direction designated "correct" by the judges.

## RESULTS

### *Structured Symbols*

1. The *t* test was used to compare the number of structured symbols sorted in agreement with the judges' ratings with chance expectancies. For all age-sex groups a greater number of structured symbols was correctly identified than would be expected by chance at less than the .01 level of significance (Table 1).

2. An analysis of variance was used to test for age and sex differences (Table 2). No significant differences were found between age groups in their identification of structured symbols.

TABLE 1

VALUES OF *t* RESULTING FROM THE COMPARISON OF THE NUMBER OF STRUCTURED SYMBOLS IDENTIFIED CORRECTLY WITH CHANCE EXPECTANCY

| Grade | Fourth | Ninth | College |
|---|---|---|---|
| Males | 3.09** | 5.67*** | 6.29*** |
| Females | 5.36*** | 4.16*** | 5.80*** |
| Both sexes | 5.77*** | 6.95*** | 8.66*** |

Note.—$n = 20$ for each subgroup.
** $p < .01$.
*** $p < .001$.

TABLE 2

STRUCTURED SYMBOLS

Correct responses
($n = 20$)

| | Male | | Female | |
|---|---|---|---|---|
| Grade | *M* | *SD* | *M* | *SD* |
| Fourth | 11.95 | 2.82 | 12.65 | 2.21 |
| Ninth | 12.95 | 2.33 | 12.3 | 2.47 |
| College | 13.55 | 2.52 | 13.4 | 2.62 |

Analysis by age and sex
($N = 120$)

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Sex | 1 | .02 | .003 |
| Age | 2 | 14.72 | 2.35 |
| Age × sex | 2 | 4.67 | .74 |
| Error | 114 | 6.27 | |

3. Reference to Table 2 indicates that male and female subjects did not differ significantly in their identification of structured symbols.

4. The significance of the subjects' agreement on the sorting of each structured symbol was tested by use of the binomial distribution. The results of these tests are presented in Table 3.

### *Texture Symbols*

1. When the average number of texture symbols identified correctly by individuals within each age-sex group was compared with chance expectancies by the use of the *t* test, it was found that the ninth-grade and college groups identified significantly more texture symbols correctly than would be expected by chance. The responses of the fourth-grade subjects did not differ from chance except as a result of a reduction in the standard error of the mean in the computation over the total fourth-grade sample (Table 4).

2. An analysis of variance was used to test for age and sex differences (Table 5). Age differences were significant at less than the .05 level. Individual comparisons were made

TABLE 3

BINOMIAL PROBABILITY OF AGREEMENT ON THE SORTING OF STRUCTURED AND UNSTRUCTURED SYMBOLS

| Structured symbols ($n$ = 40 for each grade level) | | | | Unstructured symbols ($n$ = 40 for each grade level) | | | |
|---|---|---|---|---|---|---|---|
| Symbol | Fourth | Ninth | College | Symbol | Fourth | Ninth | College |
| 1 | * | *** | ** | 21 | | ** | ** |
| 2 | *** | *** | *** | 22 | | ** | * |
| 3 | ** | *** | *** | 23 | | *** | ** |
| 4 | *** | | *** | 24 | | | *** |
| 5 | | | | 25 | | * | |
| 6 | | | † | 26 | | † | |
| 7 | | ** | *** | 27 | | ** | |
| 8 | *** | *** | *** | 28 | | | * |
| 9 | | | | 29 | * | *** | *** |
| 10 | | | | 30 | | | * |
| 11 | | | * | 31 | | | |
| 12 | | * | | 32 | *** | * | *** |
| 13 | | | ** | 33 | | | |
| 14 | | *** | *** | 34 | | *** | ** |
| 15 | * | * | ** | 35 | | *** | ** |
| 16 | *** | *** | ** | 36 | | *** | *** |
| 17 | *** | *** | *** | 37 | | | * |
| 18 | | † | | 38 | ** | *** | * |
| 19 | | | | 39 | | ** | * |
| 20 | | | | 40 | | | |

* $p < .05$.
** $p < .01$.
*** $p < .001$.
† $p < .05$ in opposition to judges' ratings.

using the studentized range distribution (Dixon & Massey, 1957). None of these comparisons reached statistical significance.

3. Significantly more females than males identified the texture symbols in agreement with the judges (Table 5).

4. The subjects' agreement on the sorting of each texture symbol was tested by the binomial distribution. The results are presented in Table 3.

TABLE 4

VALUES OF $t$ RESULTING FROM THE COMPARISON OF THE NUMBER OF TEXTURE SYMBOLS IDENTIFIED CORRECTLY WITH CHANCE EXPECTANCY

| Grade | Fourth | Ninth | College |
|---|---|---|---|
| Males | 1.28 | 4.43*** | 2.63* |
| Females | 1.89 | 6.83*** | 4.45*** |
| Both sexes | 2.29* | 7.79*** | 4.98*** |

Note.—$n$ = 20 for each subgroup.
* $p < .05$.
*** $p < .001$.

## DISCUSSION

The present data are consistent with the results of earlier studies by Starer (1955), Jones (1956), Winter and Prescott (1957), and Moos and Mussen (1959), which support predictions from Freudian theory that elongated and pointed objects would be sorted into a male category, and rounded or containing objects into a female category. In the present study the subjects of both sexes in all age groups identified the structured symbols in a manner consistent with the Freudian hypothesis. The present data are also consonant with the psychoanalytic observations of Fenichel (1945), and studies by Conn and Kanner (1947), and Brown (1958), which indicated that anatomical sex differences are recognized by boys and girls as early as ages four or five.

In contrast to the present results, Levy's (1954) fifth-grade subjects did not identify his structured symbols better than chance.

Three differences in procedure may account for the conflicting results, although these same procedural differences do not seem to be important with adults (Stennett & Thurlow, 1958). Levy's subjects matched the symbols to names rather than to nouns at an age when it is found that names are not often used to distinguish boys from girls (Conn & Kanner, 1947). Levy also employed a group procedure which may have caused difficulty in the detection of those children who, from poor motivation or lack of understanding, paired the names and symbols randomly. Finally, the population sampled in the present study may differ from that of Levy's in socioeconomic background and intellectual level.

Jones (1956) found that males identify more sexual symbols correctly than do females. Jones's findings may have resulted from an emphasis upon the sexual significance of the symbols which seemed to be contained in his instructions. Adult males may more readily respond to "sexual" symbols than adult females in the presence of a male experimenter. In the present investigation, in which no sex differences were found, the sexual nature of the symbols was not stressed, nor seemingly recognized.

TABLE 5

TEXTURE SYMBOLS

Correct responses ($n = 20$)

| | Male | | Female | |
|---|---|---|---|---|
| Grade | *M* | *SD* | *M* | *SD* |
| Fourth | 10.9 | 3.14 | 11.5 | 3.55 |
| Ninth | 12.6 | 2.62 | 13.9 | 2.55 |
| College | 12. | 3.40 | 13.8 | 3.82 |

Analysis by age and sex ($N = 120$)

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Sex | 1 | 45.64 | 4.41* |
| Age | 2 | 48.10 | 4.65* |
| Age × sex | 2 | 3.63 | .35 |
| Error | 114 | 10.34 | |

* $p < .05$.

As opposed to the responses to the structured symbols, the adolescent and adult subjects identified more unstructured symbols correctly than the prepubescent subjects, and the female subjects sorted more unstructured symbols correctly than the male subjects. The discrepancy between the data from the two types of symbols may be understood by assuming that each group of symbols refers to different aspects of sexuality. The essential bond between the symbol and the symbolized in the Freudian hypothesis is shape. The textures used, however, are relatively shapeless. This author assumes that the textures symbolize that aspect of masculinity or femininity which is culturally rather than anatomically defined, i.e., the cultural sex-role.

In the present study the subjects of all age levels identified the all black card (S-29; symbol numbers refer to Table 3) as male, and a light grey (S-38), and a soft textured card (S-32) as female. The subtle variations in texture were missed by the fourth-grade subjects which resulted in a statistically significant age difference in the number of texture symbols correctly identified. If the initial assumption is granted, one could conclude that fourth-grade children are, at some level, aware of the grosser aspects of cultural sex-roles, as indicated by Rabban (1950) and Brown (1958), but not of the subtleties which seem to be learned later.

The ninth-grade and college subjects sorted four texture symbols which were quite dark, rough, and blotchy as masculine (S-21, 22, 23, 29) and six texture symbols (S-32, 34, 35, 36, 38, 39) which were generally light in tone, airy, fluffy, and somewhat fragile in appearance as feminine. The symbols which were not consistently sorted as either masculine or feminine (e.g., S-31, 33, 40) appeared to have the characteristics mentioned above to a lesser extent or in some mixture.

Female subjects sorted more texture symbols correctly than male subjects and the difference tended to increase with age. These results suggest that the girls in the present sample were more acutely aware of the cultural sex-role stereotype than were the boys. In contrast, Lynn (1959) postulates that a young girl is somewhat confused by the lati-

tude allowed her in her sex-role behavior, and that a boy's sex-role is quite clearly defined. The apparent conflict between the views may be a function of the present writer's sample, i.e., more latitude in sex-appropriate behavior may be allowed the male in a university setting than in the modal population which may be reflected in the sorting of texture symbols.

The author has assumed that the textures used in the present study symbolize cultural sex-role expectancies, whereas the structured symbols refer only to anatomical sexuality. Post hoc examination of the sorting of each structured symbol compels one to reconsider the assumption of the present and earlier writers that structured symbols have the genitals as their sole referent. Subjects of all ages agreed that elongated symbols should be sorted in the male pile although elongatedness alone was not sufficient. The pointedness of the symbols seemed to be neither a necessary nor a sufficient element. The three symbols (S-15, 16, 17) which were sorted consistently as female by all age groups were curved in shape, but not all curved figures were similarly sorted.

It was informally observed that many symbols seemed to provide two sets of cues, one set referring to anatomical sexuality, and the other to culturally defined sexuality. The sex-role cues sometimes were consonant with the anatomical cues, in which case the symbol was sorted consistently as either male or female, e.g., an elongated and pointed symbol (S-2), seen by many as a glider, was sorted consistently into the male category. Other symbols, which were sorted irregularly, contained conflicting sex-role and shape cues (S-10, 18, 19, 20), e.g., one symbol (S-20) composed of three concentric circles was rated by judges as female, probably because of its roundness, but many subjects sorted this symbol into the male pile while informally commenting on its resemblance to a target. It is tentatively suggested on the basis of the former observations that when the anatomical and cultural cues of a structured symbol are consonant, the sorting of the symbol will be consistent. If these cues are dissonant, the response to it will be confused. There was some indication that the same confusion was contained in some of the texture symbols. For example, one dark texture symbol (S-26) composed of an entanglement of curved lines was rated as masculine by the judges, but the ninth-grade subjects agreed it should be sorted as feminine.

Although both the unstructured and structured symbols have been discussed as if the sex-role and Freudian referents could be isolated, a cultural-biological symbol dichotomy probably does not exist. Culturally, sex-role prescriptions are sometimes accepted as real biological differences (Newcomb, 1950), and psychologically, the distinction appears more theoretical than real (Colley, 1959). A dynamic interaction between the cultural and anatomical elements of a symbol might be postulated from a psychoanalytic point of view. For example, the conscious recognition of the genital referent of a structured symbol may be defended against by responding to its consonant sex-role cues, its dissonant sex-role cues, or by denying that it has any meaning. The defensive reaction to a symbol is not recorded by a nomothetic approach which tabulates only "correct" responses. The Freudian hypothesis is thus tested without sufficient regard for Freudian theory.

A further criticism of the present and earlier studies involves the communality in the cultural background of the judges and subjects. The lack of judge-subject independence would spuriously increase the number of symbols correctly identified on the basis of common cultural cues rather than because of the subjects' recognition of the anatomical meaning of the symbols. The between-subject agreement on the sorting of a symbol indicates only that the symbol had some common meaning, but does not specify if the meaning is cultural or anatomical.[3]

The interpretations of the results which have been presented are untested and are offered for their heuristic value. If it were established that "universal symbols" had cul-

[3] The studies of Barker (1957), and Schonbar and Davitz (1960) have come to the attention of the author after submission of this manuscript. Their results indicate that both adult and child subjects sort symbols with obvious cultural significance according to the cultural cues, even if these cues are at variance with the anatomical value of the symbols.

tural components, then more attention to the cultural or subcultural significance of the symbols would be in order in their interpretation. An awareness of the possible multiple dimensions of a symbol (e.g., shading and shape), and multiple determination of the response to a symbol (cultural, anatomical, and psychological) would allow greater interpretive flexibility.

## SUMMARY

Structured and unstructured sexual symbols were administered to 40 fourth-grade, 40 ninth-grade, and 40 college subjects. The structured symbols were sorted in agreement with the judges' ratings at all age levels and by both sexes. The unstructured symbols were identified in a manner consistent with judges' ratings by the adolescent and college groups, but not by the preadolescent group. A sex difference was found indicating that female subjects identified more unstructured symbols consistent with the judges' ratings than male subjects. A discussion was presented concerning the possible multiple determination of the responses to sexual symbols, the multiple dimensions of the symbols themselves, and the relation of the results to the Freudian hypothesis concerning sexual symbolism.

## REFERENCES

BARKER, E. D. The perception of sexual symbolism. Unpublished doctoral dissertation, Columbia University, 1957.

BROWN, D. G. Sex-role development in a changing culture. *Psychol. Bull.*, 1958, **55**, 232–242.

COLLEY, T. The nature and origins of psychological sexual identity. *Psychol. Rev.*, 1959, **66**, 165–177.

CONN, J. H., & KANNER, L. Children's awareness of sex differences. *J. Child Psychiat.*, 1947, **1**, 3–57.

DIXON, W., & MASSEY, F. *Introduction to statistical analysis.* New York: McGraw-Hill, 1957.

ERIKSON, E. H. *Childhood and society.* New York: Norton, 1950.

FENICHEL, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

FREUD, S. *The basic writings of Sigmund Freud.* New York: Modern Library, 1938.

JONES, A. Sexual symbolism and the variables of sex and personality integration. *J. abnorm. soc Psychol.*, 1956, **53**, 187–190.

LEVY, L. H. Sexual symbolism: A validity study. *J. consult. Psychol.*, 1954, **18**, 43–46.

LYNN, D. B. A note on sexual differences in the development of masculine or feminine identification. *Psychol. Rev.*, 1959, **66**, 126–135.

MOOS, R., & MUSSEN, P. Sexual symbolism, personality integration, and intellectual functioning. *J. consult. Psychol.*, 1959, **23**, 521–523.

NEWCOMB, T. M. *Social psychology.* New York: Dryden, 1950.

RABBAN, M. Sex-role identification in young children in two diverse social groups. *Genet. psychol. Monogr.*, 1950, **42**, 81–158.

RABIN, A. I. A contribution to the "meaning" of Rorschach's inkblots via the semantic differential. *J. consult. Psychol.*, 1959, **23**, 368–372.

SCHONBAR, R. A., & DAVITZ, J. R. The connotative meaning of sexual symbols. *J. consult. Psychol.*, 1960, **24**, 483–487.

STARER, E. Cultural symbolism: A validity study. *J. consult. Psychol.*, 1955, **19**, 453–454.

STENNETT, R. G., & THURLOW, M. Cultural symbolism: The age variable. *J. consult. Psychol.*, 1958, **22**, 496.

WINTER, W. D., & PRESCOTT, J. W. A cross validation of Starer's test of cultural symbolism. *J. consult. Psychol.*, 1957, **21**, 22.

(Received November 25, 1960)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 50–55

# THE PORTEUS MAZE TEST AS A PREDICTOR OF FUNCTIONING ABILITIES OF RETARDED CHILDREN[1]

ROBERT A. DENTLER[2] AND BERNARD MACKLER

*Bureau of Child Research, University of Kansas*

Assessment of the learning capabilities of presumably retarded children is a difficult operation. As Masland, Sarason, and Gladwin (1958), Clarke and Clarke (1958), and Haggard (1954) have shown, instruments such as the Stanford-Binet and the Wechsler-Bellevue may prove inadequate. The range of what these tests measure is relatively restricted to academic and verbal criteria. They tend to ignore cultural, motivational, and situational differences crucial to test performance. The most widely used tests also ignore the context within which particular functioning abilities are relevant.

In this report, we want to emphasize the pertinence in many contexts, and certainly the pertinence within institutions, of assessing specific functioning abilities as distinguished from measures of general intelligence. Following Baldwin (McClelland, Baldwin, Bronfenbrenner, & Stodtbeck, 1958), we have defined ability as "a characteristic of the person which permits him to behave adaptively." As Baldwin remarks, in trying to describe specific abilities, we are trying to identify systems of adaptive behavior. Moreover, ability cannot be considered to be unidimensional, as there may be a variety of ways of accomplishing the same result.

In the case of institutionalized retarded children, tests of functioning abilities should allow us to estimate the individual child's present level of ability to solve abstract problems, to comprehend the communications of others and to express himself, and to function in social situations. Measurement of relative levels of ability in these areas has greater practical utility than measurement of the extent to which general intelligence deviates from noninstitutional norms.

This report presents findings obtained in an effort to assess the value of one measure of intelligence, the Porteus Maze Test, as a predictor of three types of functioning abilities known to be pertinent to life in an institution: namely, language ability, perception of physical surroundings, and social maturity.

As Masland, Sarason, and Gladwin (1958) suggest, the Porteus Maze Test seems from past research to predict such abstract qualitative attributes as alertness, foresight, and ability to plan. It is a test that children can feel comfortable working with. The subjects in the cottage group studied in this project, boys ranging in age from 6 to 12, had with few exceptions tried and failed to remain in public schools. All had been subjected to a barrage of clinical and academic tests prior to entry into the institution and had been found lacking. Our concern was to apply tests that promised the least likelihood of tapping reservoirs of disposition to defeat or failure.

The Porteus contains no built-in indicators of language ability, perceptual awareness of surroundings, or social maturity. Unlike the intelligence tests most commonly used on retarded children, the Porteus is nonverbal and abstract. Of course, verbal instructions are involved, but they are extremely minimal. By abstract, we mean that if the Porteus measures qualities relevant to practical functioning, it does so only indirectly.

We defined the characteristic measured by the Porteus as the individual level of ability to solve abstract problems denoted by a maze of lines on pieces of paper which were not intentionally representative of or analogous to the other types of behavior

[1] This research was supported in part by United States Public Health Service Grant Number OM-111, and in part by a grant from the University of Kansas. The authors acknowledge the helpful reading and suggestions of R. L. Schiefelbusch, E. Gordon Collister, Margaret C. Byrne, Harris Winitz, and Joseph Spradlin.

[2] Now at Dartmouth College.

being assessed. Mental age (MA) on the Porteus was thus viewed as an index of an individual's ability to solve an abstract intellectual problem.

The task was to determine whether MA on the Porteus was predictive of three types of functioning ability and to assess at least one aspect of the reliability of the Porteus test as well. It is clear that the answer to the question of whether the Porteus is valid with regard to content and construct will depend upon the psychological insights of qualified and expert judgment. But the ultimate, the basic, value of the Porteus is its "functional" validity. The question here is, how well does it work in situations for which it is intended? How well does it select, differentiate or predict? Before considering the predictive power of the Porteus, we shall describe our measures of the three functioning abilities.

## MEASURES OF ABILITIES

### *Environmental Map Tests*

Two experimental map tests were developed to estimate accuracy of immediate surround. The first consisted of a two dimensional map of the floor of the institutional cottage in which the 29 subjects were housed. Its cartographic features were simple. An architectural floor plan was drawn, but complex features were removed. Lines for windows, doors, and the like, were removed. Some elements of the cottage were emphasized by representational overlay, such as the television set in the lounge, the aide's desk, the front door, and two bathrooms, and a few beds.

Each subject was presented with this Cottage Map and was asked to locate in the following order: entry, aide's area, recreation room, lounge or television room, bathroom, and his own bedroom. Finally, the subject was asked, "What room do you like most?" [3]

A second map, the Grounds Map, was a simplified two dimensional drawing of the institutional plant. An arrow was inset to identify the home cottage. Paths were drawn between those five buildings the subjects were observed to utilize most frequently. The buildings were emphasized by three dimensional insets and by enlarging their size relative to the scale of the map. The procedure was to have the subjects indicate the path to each building rather than to identify the building by its appearance. Subjects were asked how they would get to each of the following structures: dining room, gymnasium, hospital, school, and administration building.

The two map tests were significantly intercorrelated ($r = .70$, $p < .01$), permitting combination.

[3] About two thirds chose their own bedrooms.

### *The Language Test*

Ability to comprehend the communication of others and ability to verbalize are highly pertinent functioning abilities that require close assessment in the case of retarded children. The Parsons Language Sample, (Spradlin, 1960) was developed for this purpose. Using concepts of verbal behavior proposed by Skinner, Spradlin [4] devised a series of seven subtests based on 121 items that constitute tasks presented to subjects as follows.

The first subtest, Tact, requires the subject to name a series of objects and pictures. The examiner holds the object or picture in view of the child and asks, "What is it?" An appropriate vocal response is recorded for each item if the subject gives one of a number of previously specified responses. Items in the second subtest, Echoic, require the subject to repeat a series of words, phrases, or sentences, an appropriate response being a complete repetition of the vocal stimulus. The Intraverbal subtest requires the subject to complete sentences. An appropriate vocal response is determined by the part of the sentence presented by the examiner. An example is: "The flag is red, white and ————."

Subtest four, Comprehension, requires the subject to follow simple directions. Some directions are given with speech only, some directions are given with gestures only, and some are given with a combination of speech and gesture. All items in the subtest require the child to perform some specific motor act. Echoic Gestures require that the subject mimic a series of motor acts first demonstrated by the examiner.

In the sixth subtest, Intraverbal Gesture, the subject is asked questions about the function of objects. This subtest is so designed that both vocal and gestural responses are highly probable. Neither a vocal nor gestural response precludes the possibility of use of the other. The gestures elicited by these items are usually specific in nature and tend to demonstrate the function of the object. For example, if the child is asked, "What do you do with a key?" he often makes a motion with his hand as if he were twisting a key to unlock the door.

The Mand subtest consists of a series of seven items which are designed to elicit demands, requests, or questions from the subject. As in the Intraverbal Gesture subtest, an appropriate response may be a gesture and/or a vocal response. The subtests are classified as either vocal or nonvocal in the tables. The three subtests, Tact, Echoic, and Intraverbal, demand vocal responses exclusively. Thus the single measure obtained on each of the three tests is classified as vocal. The Echoic Gesture and Comprehension subtests demand nonvocal responses exclusively and the measure obtained on each subtest is classified as nonvocal. Both the Mand and Intra-

[4] The authors are grateful for the opportunity extended by Joseph Spradlin to employ his Language Sample and for his assistance in supervising the testing.

verbal Gesture subtests yield vocal and nonvocal scores.

For 40 retarded children in the same institution, the average test-retest reliability for the subtests was $r_s$ average = .90 (Spearman). To assess concurrent validity, Spradlin (1960) obtained aide rankings of retarded children on criteria of vocal and nonvocal communication ability. Rankings on vocal ability for 171 subjects were correlated with rankings on the pertinent portions of the Language Sample. The $r_s$ average for nine groups was .61. The rank correlation for 160 subjects on the nonvocal subtests with aide rankings, $r_s$ average for 10 groups was .60.

### *The Social Maturity Test*

Doll's Vineland Social Maturity Scale (1953) is a serviceable instrument for appraising an individual's social competence. Its emphasis is upon ability to perform *independently* rather than ability to adjust socially to other persons. In terms of socialization, it bears most directly upon independence training.

To assess this type of social maturity more directly, the authors devised a laboratory test based on items like those used in the Vineland interview schedule.[5] Direct assessment was deemed important in general for more reliable measurement and in particular for assessment of children entering an institution, whose parents or guardians would be responding to the Vineland schedule under volatile conditions.

A pool of 26 items ranging in difficulty from ability to unwrap candy to ability to write name with a pencil, each modeled upon Vineland items, was cast into an inventory. Subjects were tested individually in a laboratory under conditions that simulated semistructured play therapy. The tester checked success or failure on each item on the schedule as it was attempted by the child. Test-retest reliability for the 26 items was $r = .98$, with a 2-week interval between tests.

To improve the internal consistency of the instrument, the 11 items from the pool of 26 which reflected test-item intercorrelations ranging from .70 to .30 were selected to form the index.

This index of social maturity correlated .90 with standard scores from the Vineland test administered earlier by institution staff members to subject's parents.

[5] The authors are grateful for the assistance extended by Gerald Siegel in the creation and administration of this test.

### Validation of Ability Measures

As stated earlier in the paper, ability was defined in terms of adaptive characteristics. The investigators undertook correlational analysis (Pearson) with the four tests, Porteus, Maps, Language, and the Social Maturity, and measures of behavior and peer and aide social status (Dentler & Mackler, 1960).

Measures of behavior consisted of 3 weeks observation and recordings of same by the investigators. These records were transcribed by a trained recorder and acts initiated, both friendly and unfriendly, were noted.

To assess status among peers, each child's photograph was mounted in a portable board. The pictures were randomly assigned in six rows of five columns. Each child was seen individually and the questions about play, work, and identification were asked.

Aides responded to a questionnaire that consisted of questions parallel to the children's including aide preferences for play and work, and those children who responded to cottage rules and regulations.

As an indication of the association between the ability tests and behavior and sociometric status, some of the significant correlations are given. For the same 29 boys, Maturity vs. Initiates Friendly and Unfriendly Interaction = .55 ($p < .01$).

All four tests were significantly associated with sociometric peer choice status, Porteus = .45 ($p < .01$), Maps = .31 ($p < .05$), Lan-

TABLE 1

MEAN, STANDARD DEVIATION, AND RANGE OF SCORES ON MEASURES OF INTELLIGENCE AND FUNCTIONING ABILITIES

| Scores | CA | MA Porteus | MA Porteus revised | Maps | Language | Social Maturity |
|---|---|---|---|---|---|---|
| Mean | 9.61 | 5.46 | 4.38 | 5.03 | 59.24 | 5.07 |
| Standard deviation | 1.91 | 3.13 | 3.50 | 2.04 | 28.16 | 5.24 |
| Range | 6.67 to 12.75 | 0 to 15.5 | 0 to 10 | 2 to 9 | 3 to 96 | −11 to +11 |
| Maximum possible range | | | | 0 to 9 | 0 to 121 | −11 to +11 |

TABLE 2

REGRESSION COEFFICIENTS OF FUNCTIONING ABILITIES (y) ON PORTEUS MAZE MA

| Functioning abilities | Porteus | Porteus revised |
|---|---|---|
| Map Test | .66** | .59** |
| Language Test | .71** | .62** |
| Maturity Index | .64** | .75** |

Note.—$N = 29$.
** $p < .005$.

guage = .50 ($p < .01$), and Maturity = .48 ($p < .01$).

Correlation of ability tests with aide responses revealed the following significant associations: aide play preferences and Maps = .41, ($p < .01$), and Social Maturity = .48 ($p < .01$); conformity with aide regulations and Porteus = .29 ($p < .05$) and Language = .32 ($p < .05$). The above correlations are cited as indicative of the "external" validity of the ability measures.

## PROCEDURE

The four tests were administered to the 29 members of one cottage group. No child had been institutionalized prior to entry, and at the time of testing no subject had resided in the cottage longer than 2 months or less than 1. The children had been admitted to the institution and tentatively diagnosed as educable retardates. None had been placed on drugs prior to testing.

The first examiner had engaged in nonparticipant observation in the cottage for 3 weeks prior to testing and was a familiar figure to all the children. He administered the Porteus and the Map tests in one of the cottage bedrooms, the subjects having been invited casually to participate. The Cottage Map was administered first, each child being asked to point to the correct locations on the map. The Grounds Map was presented next, subjects being asked to point to the paths they would take to get to each of the five buildings. Finally, the Porteus Mazes were administered, following procedures specified in the Porteus manual (1959).

The Porteus Revision was administered by staff psychologists in the Psychological Division of the Institution, within one month after the first administration described above. The Language Test and the Maturity Index were administered by staff members of the Parsons Hospital Research Unit within the unit's laboratories. Examiners for both these tests were *not* familiar with the children.

## RESULTS

Table 1 reports means, standard deviations, and ranges of scores for the four tests of abilities. Lacking bases for conversion to standard scores on three tests, the data were transformed into $T$ scale equivalents for purposes of correlation.

The correlation between the Porteus and the Porteus Revision was $r = .85$. Several considerations lead us to interpret this as evidence of satisfactory stability-reliability. First, this was not a simple retest. Different mazes, different testers, and different testing conditions characterized the procedure. Moreover, some subjects who were able to solve maze problems on the first administration, under a tester who was very familiar with them and in the unthreatening setting of a cottage bedroom, were unable to perform at all on the administration of the second test, given by unfamiliar testers.

Table 2 indicates that the second, revised, test serves about as well as the first as a predictor of the three functioning abilities.[6] The regression coefficients support the interpretation that the Porteus Maze Test is an adequate predictor of three pertinent functioning abilities, none of which is a priori a correlate of abstract problem solving ability. Chronological age was of course positively associated with MA. The rather large CA range of from 6–8 years to 12–9 years affected the correlations. When CA was held constant, however, each regression coefficient remained significant at the .025 level or better (Table 3).

The intracorrelation of tests by subjects is also of interest. The subjects were ranked on the Porteus, Maps, Language, and Social

[6] Since $T$ scale equivalents were used, incidentally, the regression coefficients are exactly equal to correlation coefficients between the variables. Cf., Allen E. Edwards (1954).

TABLE 3

INTERCORRELATIONS ($r$) OF FUNCTIONING ABILITIES WHEN CHRONOLOGICAL AGE IS HELD CONSTANT

| Functioning abilities | Porteus | Porteus revised |
|---|---|---|
| Map Test | .53* | .44* |
| Language Test | .48* | .43* |
| Maturity Index | .58* | .70* |

Note.—$N = 29$.
* $p < .025$.

TABLE 4

DISCREPANCIES BETWEEN RANKED SCORES

| Test | *N* | Number cases showing discrepancy between ranks with Porteus | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Below −15 | −15 to −11 | −10 to −6 | −5 to +5 | 5 to 10 | 11 to 15 | 16 to 20 |
| Maps | 29 | 0 | 1 | 5 | 17 | 5 | 1 | 0 |
| Social Maturity | 29 | 0 | 3 | 1 | 18 | 6 | 1 | 0 |
| Language | 29 | 1 | 1 | 2 | 19 | 3 | 3 | 0 |

Maturity tests, and Kendall's coefficient of concordance was calculated. For the four sets of scores, $W = .80$, ($p < .001$) which indicates a very high degree of concordance among the scores taken by subject.

Table 4 displays in more concrete fashion the trend of discrepancies between the Porteus and the three tests of functioning abilities. On the average, two thirds of the subjects reflected identical or virtually identical ranks. Discrepancies in ability scores were *not* associated with differences in CA.

As a further check, the examiners attempted to answer another question. Does the general mental level of the child influence the amount of intertest discrepancies? Analysis of discrepancies by mental level on the Porteus revealed a complete *lack* of effect from mental level. That is, discrepancies do not vary in magnitude from MA level to MA level.

## DISCUSSION

The Porteus is a fairly stable and adequate predictor of language ability, awareness of environmental surroundings, and social maturity. This simple, brief, test designed to measure abstract problem solving ability is useful in making environmentally pertinent distinctions between retarded children. The broad utility of the Porteus Maze Test is once again confirmed.

Investigating the discrepancy between scores (Table 4) more closely, we found 11 scores disagreed by at least 11 ranks with rank on the Porteus. Seven of the 29 children accounted for these discrepancies, however. Three of the seven reflected discrepancies between Porteus and two of the three functioning abilities.

In order to consider discrepancies analytically we viewed each individual's score on the Porteus as an anchor point. Among the seven individuals with notable differences between their ranks on the Porteus and their ranks on the other tests, we found two crude types: subjects whose language, environmental, and social maturity performances were *below*, and subjects whose performances were *above*, the rank obtained on the Porteus. Discrepancies may serve as crude profiles. The baseline for profile analysis here is not an "outside" population, of course, but the child's peers within the institution. Ideally, this baseline would include a large sample of all currently institutionalized children at a particular institution.

To illustrate this use of discrepancies, consider one of the boys in this sample who appears to be functioning *below* his Porteus MA level. He ranked third on MA but fifteenth on the Maps and fifteenth on Language. After his third month in the institution, this boy was diagnosed as capable of effective performance in special education classes in public schools and was released from the institution.

Properly considered, each test assessed the ability of each child to perform selected tasks under very special *social* conditions. The examiners were adults and the tests was administered under the limits of a two-person interaction situation. A more valid assessment of functioning abilities would include samples of behavior among peers and in more complex or variegated social settings. By viewing the Porteus Maze Test as a test of a particular kind of functioning and by relating it to tests of other types of behavior, two things were learned. For three out of four subjects, the Porteus Maze was an excellent predictor of level of functioning. For other

subjects, the discrepancies between the Porteus and the other tests were found to relate meaningfully to the behavior of the children within the institutional cottage.

The aim here has been to suggest very tentatively the interpretive utility of the evaluation of discrepancies. If a greater variety of measures of functioning abilities was used, this approach could be extended. Academic skills, subtests of particular language abilities, and motor skills, perhaps, could be added. The additional discrepancies between tests could facilitate rehabilitation and training planning for individual children.

## SUMMARY

To explore the adequacy of the Porteus Maze Test as a predictor of three types of functioning abilities of retarded children, tests of language ability, cognitive awareness of environmental surrounds, and social maturity were administered to a cottage group of 29 boys, newly admitted to a training institution, who ranged in age from 6 to 12 years.

The Porteus was selected as a general measure of learning capability because of its abstract, unthreatening characteristics and because of ease and economy of administration. Scores from two Map tests—maze type sketches of the children's new cottage and of the hospital grounds—were combined as representing the ability of the child to identify representational symbols of environmental landmarks and the paths between them.

The test of language ability was an instrument previously developed by psychologists employing Skinnerian categories of verbal behavior and tentatively established as sufficiently reliable and valid. The authors devised an index of social maturity by converting items from the Vineland Social Maturity Scale into tasks performed under laboratory conditions. The index was found to be highly stable and to associate highly with measures of interpersonal behavior.

The Porteus Maze Test was found to be sufficiently stable and adequately predictive of Map, Language, and Maturity scores. Discrepancies between the four tests were explored discursively as indicators of diagnostic assessment, with the suggestion that this approach might prove valuable for institutional planning of individual regimens.

## REFERENCES

CLARKE, A. M., & CLARKE, A. D. B. *Mental deficiency*. Glencoe, Ill.: Free Press, 1958.

DENTLER, R. A., & MACKLER, B. Mentally retarded children and the sociometry of institutionalization. Unpublished manuscript, Bureau of Child Research, University of Kansas, 1960.

DOLL, E. A. *The measurement of social competence*. Minneapolis: Educational Test Bureau, 1953.

EDWARDS, A. E. *Statistical methods for the behavioral sciences*. New York: Rinehart, 1954, Pp. 161.

HAGGARD, E. A. Socio-status and intelligence. *Genet. psychol. Monogr.* 1954, **49**, 141–186.

MASLAND, R. L., SARASON, S. B., & GLADWIN, T. *Human subnormality*. New York: Basic Books, 1958.

MCCLELLAND, D. C., BALDWIN, A. L., BRONFENBRENNER, U., & STODTBECK, F. L. *Talent and society*. Princeton: Van Nostrand, 1958.

PORTEUS, S. D. *The maze test and clinical psychology*. Palo Alto: Pacific Books, 1959.

SPRADLIN, J. E. The Parsons Language Sample. Unpublished manuscript, Parsons State Hospital, Parsons, Kansas, 1960.

(Received December 7, 1960)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 56–60

# PREMORBID BEHAVIOR AND PROGNOSIS IN FEMALE SCHIZOPHRENIC PATIENTS[1]

AMERIGO FARINA,[2] NORMAN GARMEZY,[3] MARIAN ZALUSKY, AND JOSEPH BECKER[4]
*Duke University*

Considerable research effort has been expended in determining those characteristics of schizophrenic patients which have prognostic significance. This search for prognostic indicators has a practical purpose, since knowledge about the course of this illness is of value both to the hospital and to the patient's family in planning for the future. Of perhaps even greater value, however, is the contribution which such research may make to a better understanding of the disease process itself by providing cues about the relevant antecedent events and personality attributes of patients which are associated with remission. The many difficulties encountered in attempts to evaluate prognostic factors have been discussed in detail by Blair (1940) and Huston and Pepernik (1958).

For male schizophrenic patients, marriage is one factor consistently reported to correlate with recovery and discharge from the hospital. Counts and Devlin (1954), Jenkins and Gurel (1959), Lindemann, Fairweather, Stone, and Smith (1959), and Mason, Tarpy, Sherman, and Haefner (1960) all report that male married patients recover more rapidly than patients who are single. This differential rate of recovery may indicate that the more mature and more adequately functioning males are more likely to marry and, because of their greater maturity, are also likely to recover more rapidly from a schizophrenic episode than single males.[5] The greater social maturity of married as compared to single males is suggested by the fact that to marry each, typically, has had to make advances to a female, court her, propose marriage and offer a home and economic support. The argument that adequacy of premorbid adjustment is related to recovery is supported by several prognostic studies employing the Phillips Premorbid Scale (Phillips, 1953). This scale quantifies premorbid adequacy in terms of social and sexual adjustment and marital status contributes to a patient's total score. Studies employing the scale, (Farina & Webb, 1956; Phillips, 1953; Seidel, 1960) all report that patients who have achieved more adequate levels of premorbid adjustment tend to recover more rapidly from a schizophrenic episode than do those with a poor premorbid history. Laboratory studies indicating more adequate functioning under stress of good as compared to poor premorbid schizophrenic patients have been summarized recently (Garmezy & Rodnick, 1959; Rodnick & Garmezy, 1957). In summary, it would appear that good premorbid adjustment is associated both with marriage and remission for male patients and that greater ego maturity is an underlying factor.

Unfortunately, prognostic studies have tended to concentrate on males and there have been relatively few studies of female schizophrenic patients. Orr, Anderson, Martin, and Philpot (1955), in studying the degree of association between a number of prognostic factors and recovery in a group of female patients, report that, although marital status was associated with recovery, length of hospital stay and education were better predictors of remission. In another study of prognostic factors involving mainly female patients, Schofield, Hathaway, Hastings, and Bell (1954) failed to find a relationship between marital status and recovery. Thus the evi-

[1] This investigation was supported by a USPHS research grant (M-629) from the National Institute of Mental Health, United States Public Health Service.

[2] Now at the University of Connecticut.
[3] Now at the University of Minnesota.
[4] Now at the University of Illinois.

[5] The greater likelihood that married (relative to single) patients will leave and remain out of the hospital does not seem to be due to great conjugal family tolerance for pathological behavior. Davis, Freeman, and Simmons (1957) have presented evidence indicating that parental families are more willing to tolerate partially remitted patients at home than are conjugal families.

dence pertinent to determining the prognostic significance of marital status for schizophrenic females seems ambiguous and apparently no data on the significance of premorbid adjustment for recovery are available.

This raises the question of whether marital status and level of premorbid adjustment also are associated with recovery from schizophrenia in the case of females. Here the prediction is less clear than with males. A constellation of characteristics such as extroversion, a capacity to establish and maintain friendships, and the ability to interact smoothly with others—all of which are aspects of adequate premorbid functioning—would appear to be prognostically favorable for both men and women. Such characteristics would seem to be inconsistent with the development of a chronic schizophrenic syndrome. For women, however, marital status may be less strongly associated with recovery from the disorder, since marriage need not imply the same degree of maturity as it does for males. In our society the woman's role in the marriage process can be a very passive one and, indeed, if she has a detached, nonargumentative disposition, she may appear more feminine and more attractive to some males. The implication of this position is that for schizophrenic females social aspects of premorbid adjustment should be more strongly associated with remission than marital status.

This study was undertaken in order to determine the prognostic significance of both marital status and level of premorbid adjustment for schizophrenic females. It was hypothesized that adequacy of premorbid adjustment would be more strongly associated with recovery for this type of patient than would marital status.

## PROCEDURE

The data for the study were obtained from records available at two North Carolina state hospitals.[6] The hospitals employ similar treatment procedures, each receives its patient population from neighboring counties, and the patients housed at each hospital seem to be of comparable ages and social classes. All first admission female patients admitted to these hospitals between July 1, 1949 and December 31, 1952 were included in this study if they were between the ages of 17 and 40 and bore a staff diagnosis of schizophrenia uncomplicated by mental deficiency or organic brain damage. Each patient was traced for a minimum of 5½ years after hospitalization, and the time spent in and out of the hospital was cumulated. In order to maximize the possibility of observing relationships between patient characteristics and recovery, a rigorous criterion for recovery was used and two extreme groups were selected. All patients who remained in the hospital no more than 6 months and who were out of the hospital at least 5 years were included in the Recovered Group. The Nonrecovered Group consisted of those patients who remained in the hospital 5 years or more and who were out of the hospital for a total period of less than 6 months. All patients whose hospital histories placed them between these two groups were excluded from the study.

After the two groups of subjects had been identified, the following data were tabulated for each subject from records obtained at the time of her admission: marital status, age, education, work history, and number and sex of younger and older siblings. All subjects were also rated for adequacy of social and sexual premorbid adjustment by means of the Phillips Premorbid Scale which was slightly modified to facilitate its use with female patients. The scale is composed of five subscales on each of which the patient is assigned a score ranging from 0 to 6. Subscale A is designed to measure adequacy of recent sexual adjustment; B, social aspects of the patient's sexual life during the adolescent period; and C, social aspects of recent sexual life. Subscales D and E are intended to quantify nonsexual aspects of premorbid interpersonal functioning, the former being concerned with past and the latter with more recent inter-

TABLE 1

CHARACTERISTICS OF RECOVERED AND NONRECOVERED GROUPS OF SUBJECTS

| Subjects | | Recovered | Nonrecovered |
|---|---|---|---|
| All subjects | *N* = | 50 | 33 |
| Age at admission | Mean | 28.8 | 29.5 |
| | *SD* | 6.93 | 5.2 |
| Education in grades completed | Mean | 8.3 | 9.8 |
| | *SD* | 3.15 | 2.96 |
| Siblings | Mean | 5.5 | 5.8 |
| | *SD* | 2.44 | 2.87 |
| Per cent working prior to admission | | 70% | 87% |
| Married, separated, and divorced subjects | *N* = | 42 | 22 |
| Age at marriage | Mean | 19.3 | 21.7 |
| | *SD* | 2.83 | 4.16 |
| Number of children | Mean | 2.6 | 1.9 |
| | *SD* | 2.07 | 1.70 |

[6] Appreciation is expressed to the administrative and professional staffs of Dorothea Dix Hospital, Raleigh, North Carolina, and John Umstead Hospital, Butner, North Carolina, for their cooperation in this study.

TABLE 2

MARITAL STATUS OF SUBJECTS AT ADMISSION

| Subjects | | Recovered | Nonrecovered |
|---|---|---|---|
| Married | Number | 39 | 16 |
| | Percentage | 78 | 49 |
| Single | Number | 8 | 11 |
| | Percentage | 16 | 33 |
| Separated or divorced | Number | 3 | 6 |
| | Percentage | 6 | 18 |

personal relationships. An individual's total score, therefore, can vary from 0, indicating excellent premorbid adjustment, to 30 for dismally poor adjustment prior to hospitalization.

In order to avoid bias in rating, a clinical psychologist experienced in the use of the scale rated each patient without knowledge of her hospital history. This was accomplished by providing the judge with only the social history portion of each patient's record. Since the social history had been obtained in each case at the time of admission to the hospital it contained no biasing information about subsequent outcome of the illness. To determine the reliability of the scale, a subgroup of 22 subjects, representing a random sample drawn from each group, was independently rerated by a second judge. The resulting Spearman rho for the total scale was .86, indicating a reasonably satisfactory degree of reliability.

For the married, divorced, or separated subjects of both groups, age at marriage and number of children were also obtained in addition to data tabulated for all subjects as described above.

## RESULTS

Table 1 compares the Recovered and Nonrecovered Groups on several variables unrelated to the major purposes of this study. It is apparent that these groups are quite similar with regard to age at admission and number of siblings. There is a trend for the Nonrecovered Group to have had more education than the Recovered Group. This difference, while not statistically significant, is consistent with previous findings for both male (Farina & Webb, 1956) and female patients (Orr, et al., 1955). The married, separated, and divorced subjects of each group are also compared on age at marriage and number of children in Table 1.

The findings concerning the marital status of the subjects in each group at the time of their first admission are presented in Table 2. It is readily apparent from this table that a greater percentage of Recovered subjects are married at admission than Nonrecovered subjects. The significance of this difference in proportion of married to nonmarried (single, separated, and divorced) patients was evaluated by means of a chi square test. The resulting chi square is 6.48 which, for one degree of freedom, is significant at better than the .05 level and suggests that married females are more likely to recover from a schizophrenic process than are single, separated, or divorced females.[7]

The analysis of scores derived from the Phillips Premorbid Scale revealed that the Nonrecovered Group has a mean scale score about three points higher than the Recovered Group. The statistical significance of this difference in scores was determined by means of a Mann-Whitney *U* test (Mann & Whitney, 1947) and the resulting *z* was 2.36, which is significant at the .05 level. It may be concluded, therefore, that relatively adequate premorbid adjustment is associated with comparatively rapid and enduring recovery from a schizophrenic episode for this sample of patients.

Thus the results indicate that both marital status and adequacy of premorbid adjust-

[7] When the samples from each of the two hospitals providing subjects are separately analyzed, the Recovered and Nonrecovered Groups from one hospital are not significantly different with regard to the proportion of married to nonmarried patients. However, the trends are in the direction shown in Table 2.

TABLE 3

RELATIVE DEGREE OF ASSOCIATION OF 13 VARIABLES WITH RECOVERY

| Variable | Rank order | Percent of additional variance accounted for |
|---|---|---|
| Subscale C of Phillips scale | 1 | 10.6 |
| Education | 2 | 5.1 |
| Subscale D of Phillips scale | 3 | 4.0 |
| Worked prior to admission | 4 | 3.1 |
| Siblings—older male | 5 | 1.8 |
| Subscale E of Phillips scale | 6 | 1.4 |
| Age at admission | 7 | 1.0 |
| Siblings—younger male | 8 | .8 |
| Siblings—older female | 9 | .5 |
| Siblings—younger female | 10 | .3 |
| Marital status | 11 | .3 |
| Subscale A of Phillips scale | 12 | .05 |
| Subscale B of Phillips scale | 13 | .02 |

ment are prognostically significant variables. However, since a correlation of .74 was found between marital status and total premorbid rating, the prognostic significance of premorbid social adjustment was evaluated holding marital status constant. This was accomplished by determining, for a homogeneous group of married patients, the relationship of Subscales D plus E to recovery. These two groups of items tap social functioning and, like the other subscales, good adjustment is indicated by low scores. The proportion of Recovered patients whose D plus E score fell below the median for both groups was 71% while the corresponding proportion of Nonrecovered patients was 25%. This difference is significant at the .05 level ($\chi^2 = 5.63$) and can be interpreted as indicating that even in a group of female patients homogeneous with regard to marital status, the more adequate the level and extent of interpersonal relationships manifested prior to the disorder, the greater the likelihood of remission.

A further attempt was made to determine the relative degree of association of marital status and premorbid functioning with recovery. A regression analysis[8] was performed using 13 variables, including marital status and the five Phillips subscales. In this analysis, the most important variable in terms of proportion of variance accounted for in the dependent measure was selected out. The process was then repeated for the remaining 12 variables and continued until only one variable remained. The rank order of these variables as determined in this way, as well as the additional proportion of variance accounted for by each is presented in Table 3. For these samples of subjects, Subscale C of the Phillips scale is most strongly associated with recovery. While the score obtained on this subscale is in part a function of marital status, adequacy of nonsexual interpersonal relationships also determines the score. Furthermore, Items D and E, which are concerned exclusively with social relationships, show a relative association with recovery. In contrast to this, marital status clearly contains little or no unique information which would aid in making a prognostic statement. In spite of its significant association with recovery, these results indicate that when the other variables preceding it are known, marital status adds virtually nothing to such a prediction. It is ranked as eleventh most important of the 13 variables, being followed only by Subscales A and B of the Phillips scale, and these three variables jointly account for less than 1% of the variance in the dependent measure. The close rankings of these three variables is understandable since Subscales A and B measure presence and stability of heterosexual relationships and such relationships, especially for women, would seem to vary with marital status.

[8] Appreciation is expressed to C. A. Boneau for his assistance with this analysis which entailed the use of an IBM 650 computer.

## DISCUSSION

The reported relationship between premorbid adjustment and recovery seems to indicate that it is a particular type of patient who is likely to develop a form of chronic schizophrenia characteristic of a large segment of our state hospital populations. As with male patients, schizophrenic women who have had difficulty establishing satisfactory interpersonal relationships and who remained withdrawn and aloof from others prior to the disorder tend not to recover in comparison to women whose premorbid interpersonal relationships were more adequate.

The obtained relationship between marital status and recovery may indicate that it is the better adjusted and more mature women who marry and who, because of their greater ego resources, recover more rapidly and for longer periods of time than do single women. However, many married subjects had shown very poor social adjustment prior to the onset of schizophrenia. Presumably the relationship between marriage and recovery could be made much more comprehensible by an analysis of such factors as the patient's motivation for marriage, the personality of the husband, and the stability of the marriage itself. For example, a woman who marries because of strong dependency needs, to seek refuge from an unhappy family life, or because of passive compliance with society's pressures for conformity may be a poorer candidate for recovery than a single woman unmotivated by such immature considerations. In contrast to these motivations, a woman who marries be-

cause of love and a desire for the responsibilities and pleasures entailed in having a family of her own may be expected to be a better prognostic risk if she develops schizophrenia. More intensive research into the marriage process itself and the motivations underlying marriage and the choice of a spouse will be necessary to determine the validity of these hypotheses.

The results of the study offer some tentative support for the hypothesis that, for female schizophrenic patients, marriage is less strongly associated with remission than is adequacy of social premorbid adjustment. Whether comparable findings would characterize male schizophrenic patients or whether there would be a stronger relationship between marriage and other aspects of social functioning can only be answered by further research. A study comparable to this one, using male patients and the same criteria for recovery and nonrecovery, is now underway.

Finally, some methodological implications may be derived from this and other studies which have been cited. In studies of prognosis in schizophrenia, it would appear to be important to separate patients by sex before analyzing remission data in view of the different meaning which such factors as marriage may have for males as compared to females. Failure to make such separations may either obscure existent relationships or result in inadequate or inaccurate generalizations. Furthermore, in studying the relationships between premorbid functioning and recovery, married and single patients should be considered separately because of the high positive correlation which exists between marital status and other aspects of premorbid adjustment.

## SUMMARY

A cohort of first admission female schizophrenic patients was identified and two groups were selected for study. One group had demonstrated a relatively rapid and enduring recovery in terms of release from a mental hospital while the other group had failed to make such a recovery. Married patients and patients whose social interpersonal relationships prior to the disorder were reasonably good tended to make a relatively rapid recovery from schizophrenia. Tentative evidence was obtained indicating that adequate interpersonal adjustment was prognostically more significant than marital status for this group of female patients. Some methodological implications of this and similar studies are discussed.

## REFERENCES

BLAIR, D. Prognosis in schizophrenia. *J. ment. Sci.*, 1940, **86**, 378–477.

COUNTS, R. M., & DEVLIN, J. P. Sexual experience as a prognostic factor in psychosis. *J. nerv. ment. Dis.*, 1954, **120**, 364–368.

DAVIS, J. A., FREEMAN, H. E., & SIMMONS, O. G. Rehospitalization and performance level among former mental patients. *Soc. Probl.*, 1957, **5**, 37–44.

FARINA, A., & WEBB, W. W. Premorbid adjustment and subsequent discharge. *J. nerv. ment. Dis.*, 1956, **124**, 612–613.

GARMEZY, N., & RODNICK, E. H. Premorbid adjustment and performance in schizophrenia: Implications for interpreting heterogeneity in schizophrenia. *J. nerv. ment. Dis.*, 1959, **129**, 450–466.

HUSTON, P. E., & PEPERNIK, M. C. Prognosis in schizophrenia. In L. Bellak (Ed.), *Schizophrenia: A review of the syndrome*. New York: Logos, 1958. Pp. 531–546.

JENKINS, R. L., & GUREL, L. Predictive factors in early release. *Ment. Hosp.*, 1959, **10**, 11–14.

LINDEMANN, J. E., FAIRWEATHER, G. W., STONE, G. B., & SMITH, R. S. The use of demographic characteristics in predicting length of neuropsychiatric hospital stay. *J. consult. Psychol.*, 1959, **23**, 85–89.

MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.

MASON, A. S., TARPY, ELEANOR E., SHERMAN, L. J., & HAEFNER, D. P. Discharges from a mental hospital in relation to social class and other variables. *AMA Arch. gen. Psychiat.*, 1960, **2**, 1–6.

ORR, W. F., ANDERSON, RUTH B., MARTIN, MARGARET P., & PHILPOT, D. F. Factors influencing discharge of female patients from a state mental hospital. *Amer. J. Psychiat.*, 1955, **111**, 576–582.

PHILLIPS, L. Case history data and prognosis in schizophrenia. *J. nerv. ment. Dis.*, 1953, **117**, 515–525.

RODNICK, E. H., & GARMEZY, N. An experimental approach to the study of motivation in schizophrenia. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957*. Lincoln: Univer. Nebraska Press, 1957. Pp. 109–183.

SCHOFIELD, W., HATHAWAY, S. R., HASTINGS, D. W., & BELL, D. M. Prognostic factors in schizophrenia. *J. consult. Psychol.*, 1954, **18**, 155–166.

SEIDEL, CLAUDENE, The relationship between Klopfer's Rorschach prognostic scale and Phillips' case history prognostic rating scale. *J. consult. Psychol.*, 1960, **24**, 46–49.

(Received December 8, 1960)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 61–64

# DEPENDENCY FANTASY AND GROUP CONFORMITY IN ULCER PATIENTS[1]

PETER WEISS[2] AND WALTER EMMERICH

*University of Colorado*

Alexander and his colleagues have hypothesized that dependency frustration is a critical factor in the etiology of peptic ulcer (Alexander, 1934, 1950; Alexander & French, 1948). According to this view, the close association during infancy between the feeding process and receipt of maternal affection results in their becoming equated symbolically. Continued frustration of interpersonal expressions of dependency later in life may lead to regressive conversion into the wish to be fed, resulting in chronic gastric stimulation, and, in some cases, ulceration. It has been observed that some ulcer patients (Primary Type) express their dependency needs directly and excessively, whereas others (Reactive Type) repress these needs and overcompensate through exaggerated ambitiousness and independence (Kapp, Rosenbaum, & Romano, 1947; Marquis, Sinnett, & Winter, 1952; Scodel, 1953; Winter, 1955). The utility of this typology depends upon the actual presence of each type within the population of ulcer patients and the relative absence of each type from a variety of nonulcer populations. The present study was concerned with one aspect of this problem, and explored the hypothesis that Primary Type ulcer patients exhibit more dependency in both their fantasied and actual social relationships than nonulcer psychosomatic and nonpsychosomatic controls.

It seemed reasonable to assume that by selecting ulcer patients from the lower socioeconomic levels, most of the sample would consist of the Primary, and not the Reactive Type. Obviously, such persons have had limited opportunities to prepare for or enter these occupations in which strong mobility strivings are highly sanctioned. It is also believed that their family backgrounds indirectly or even actively discourage the internalization of status-seeking and achievement motives (Scodel, 1953). Further, there is some direct evidence that lower socioeconomic ulcer patients tend to have Primary Type personality characteristics (Scodel, 1953; Winter, 1955).

n Succorance on the TAT (Murray, 1943) was chosen as the fantasy measure, as it clearly reflects the kinds of fantasies that have been attributed to the Primary Type ulcer patient. Actual interpersonal dependency was assessed in the group conformity situation developed by Asch (1952). The amount of conformity elicited in this situation presumably reflects the strength of the individual's need to depend upon the judgments of others. Evidence that n Succorance on the TAT and Asch's conformity measure refer to the same construct of dependency is provided by Kagan and Mussen's (1956) finding of a positive relationship between these two measures in male undergraduates.

## METHOD

### *Subjects*

The subjects were 38 male patients from the Veterans Administration Hospital, Denver, Colorado, 74% of whom had been hospitalized for 15 days or less. Their ages ranged from 25 years to 65 years, with a median of 46 years. Thirty-three were married; four were divorced, separated, or widowed; and one was single. Hospitalization in a Veterans Administration hospital was taken as presumptive evidence of lower-middle or lower-class membership. A crude check of this assumption was provided by occupational information secured from hospital admission cards. None of the subjects reported having professional, semiprofessional, or entrepreneurial positions, and 71% of the sample indicated that they were either semiskilled or skilled laborers.

The three subgroups were formed as follows: (*a*) The ulcer (experimental) group was selected from the available gastrointestinal patients who were

[1] This paper is based upon the senior author's thesis in partial fulfillment of the requirements for the MA degree, University of Colorado. The authors wish to express their appreciation to the staff of the Veterans Administration Hospital, Denver, Colorado, whose cooperation made this study possible. Special thanks go to Abby Adams for her assistance in collecting the data.

[2] Now at Olive View Sanitorium, California.

diagnosed as having peptic or duodenal ulcers uncomplicated by additional psychosomatic disorders. (b) The nonulcer psychosomatic control group were patients from the general medicine and surgery services whose diagnoses indicated that they had disorders of psychosomatic origin, such as asthma, rheumatoid arthritis, and neurodermatitis. (*c*) The nonpsychosomatic control group consisted of patients having disorders believed not to be psychosomatic, such as cancer, genito-urinary infection, and joint dislocation. The three groups were matched roughly for age, marital status, occupation, and length of current hospitalization. Since some subjects were not available for testing on both the TAT and conformity measures, the subgroup *N*s differed for each of these measures (see Table 1).

### *Assessment of Dependency Fantasy*

The TAT administration procedure was similar to that developed by McClelland, Atkinson, Clark, and Lowell (1953). Groups of 5 to 10 patients responded on a standardized form to the TAT cards projected on a screen. The TAT cards used by Kagan and Mussen (1956) to measure dependency were presented in the following order: 1, 3BM, 6BM, 7BM, 12M, 13B, 14, 18BM. The experimenter read the following instructions to the group:

> The Veterans Administration Hospital is interested in your reactions to the pictures which we are going to show. It is part of a big research program, designed to modernize the patient program here. Please do not put your name on these papers.
>
> A number of pictures will be projected on the screen before you. You will have 20 seconds to look at the picture and then 5 minutes to make up a story about it. Notice that there is one page for each picture. The same four questions are asked about each picture. They will guide you in your thinking and enable you to cover all of the elements of a plot in the time allotted. I will keep time and tell you when to go on to the next question for each story.
>
> It is important to remember that there is no right or wrong way to do this. We are interested in your reactions, so feel free to make up any kind of story you choose. Do not merely describe what you see in the picture, but rather use your imagination and tell a good story. Make the pictures come alive in your stories. Work as fast as you can in order to finish in time. If you need more space for any question, use the reverse side of the page. The staff joins me in thanking you for your interest in this survey. Are there any questions?

The standardized form consisted of the following sequence of questions, taken from McClelland et al. (1953, p. 98): (*a*) What is happening? Who are the persons? (*b*) What has led up to this situation? That is, what has happened in the past? (*c*) What is wanted? By whom? (*d*) What will happen? What will be done?

Using Kagan and Mussen's (1956) criteria, a dependency score was recorded for themes in which the hero seeks help from another person in solving a personal problem, or is disturbed over the loss of a source of love or support. The dependency fantasy score was the total number of cards containing a dependency theme, and could vary from zero to eight. In order to avoid halo effects, the protocols for each card were rated across subjects. Stories were rated without knowledge of the subjects' diagnoses or conformity scores. There was perfect agreement between two judges' independent ratings of the 40 protocols of 5 randomly chosen subjects.

Administration of the TAT always preceded the conformity test, and a patient received both tests on the same day. In order to minimize the possibility that subjects would connect the two parts of the study, different experimenters and experimental rooms were used for each part.

### *Assessment of Conformity*

Seven confederates participated in the Asch (1952) group conformity situation. When subject entered the experimental room, the confederates had occupied chairs arranged in two rows of four facing experimenter. The remaining chair was placed so that subject would be the seventh person to announce his judgment. Experimenter read the following instructions to the group:

> This is a task which involves the discrimination of the lengths of lines. You see a pair of white cards in front. The card on the left shows a single line; the card on the right has three lines differing in length. They are numbered 1, 2, and 3 in order. One of the three lines on the card on the right is equal in length to the standard line on the card on the left. You will decide in each case which is the equal line. You will state your judgment in terms of the corresponding number; either one, two, or three. There will be twelve such comparisons. Since the number of lines is few and the group is small, I shall call upon each of you in turn to announce your judgment which I shall record here on a prepared form. Please be as accurate as possible. Suppose we start at the right and proceed to the left and back again to the right in the second row.

The 12 pairs of cards were identical to those used by Asch (1952). The seven confederates were given prior instructions to respond correctly on Trials 1, 2, 5, and 11, and to give the same incorrect response on the remaining seven trials. The conformity score was the total number of yields on the seven critical trials. Upon completion of the conformity test, subject was informed of its nature, and was invited to participate as a confederate in the test for the next subject. The other confederates were recruited from the neuropsychiatric service. Confederates and subjects were never more than casually acquainted with each other prior to the conformity test.

TABLE 1

GROUP MEDIANS AND COMPARISONS ON THE DEPENDENCY FANTASY AND GROUP CONFORMITY MEASURES

| | | Dependency fantasy | | Group conformity | |
|---|---|---|---|---|---|
| | Group | *N* | Median[a] | *N* | Median[a] |
| Group medians | Ulcer (U) | 10 | 5.0 | 14 | 3.0 |
| | Nonulcer psychosomatic (P) | 8 | 4.0 | 9 | 0.5 |
| | Nonpsychosomatic (N) | 12 | 2.5 | 14 | 1.0 |
| | | | $p$[b] | | $p$[b] |
| Group comparisons | U × P | | n.s. | | <.05 |
| | U × N | | <.05 | | <.05 |
| | P × N | | <.05 | | n.s. |

[a] Rounded to the nearest multiple of .5.
[b] $p$ values are for one-tailed Mann-Whitney *U* tests corrected for tied ranks.

## RESULTS AND DISCUSSION

Both the ulcer and nonulcer psychosomatic groups wrote significantly more dependency themes on the TAT than did the nonpsychosomatic controls (see Table 1). However, although the ulcer group was more dependent than the nonulcer psychosomatics on this measure, the difference was not significant. Thus, the hypothesis with respect to dependency fantasy was only partially confirmed.

The hypothesis was more consistently supported on the Asch conformity measure. The ulcer group yielded significantly more frequently than both the nonulcer psychosomatic and nonpsychosomatic controls, and the latter two groups did not differ significantly from each other (see Table 1).

The results confirm the belief that Primary Type ulcer patients are more dependent in both fantasy and actual social behavior than nonpsychosomatics. However, since both the ulcer and nonulcer psychosomatic groups were differentiated on the fantasy measure from the nonpsychosomatics, but not from each other, it would appear that strong dependency needs are not specific to ulcer among the psychosomatic disorders. This conclusion is supported by the clinical observation that unresolved dependency occurs in a number of psychosomatic disorders (Alexander, 1950; Alexander & French, 1948), and by Streitfield's (1954) finding of no difference between ulcer and nonulcer psychosomatic patients on Rorschach and Blacky measures of oral dependency.

The fact that group conformity was greater in the ulcer than in the nonulcer psychosomatic group is consistent with the findings of Scodel (1953) and Marshall (1960), and indicates that specificity lies in the way that dependency is mediated in actual behavior, rather than in need strength per se. Whereas the ulcer group tended to express their dependency needs directly, the nonulcer psychosomatics inhibited these needs in the conformity situation. Such inhibition, together with the relatively strong dependency wishes revealed by the TAT, suggests that the nonulcer psychosomatic patient has considerable anxiety about his dependency needs. It is possible to explore this hypothesis by comparing the groups on TAT dependency themes rated for dependency-anxiety. A dependency theme was scored as anxious if its ending indicated nonfulfillment of dependency because of object loss, rejection, or punishment by the object of nurturance. This criterion was met by 34% of the themes in the ulcer group, 50% in the nonulcer psychosomatic group, and 26% in the nonpsychosomatic group.[3] These findings are consistent with the interpretation, but further research is necessary to provide a more definitive test and to determine if dependency-anxiety in nonulcer psychosomatics is accompanied by overcompensation and/or other specific defenses.

[3] Significance tests of group differences on this measure were not feasible because of the large percentage of subjects who received zero scores in conjunction with small subgroup *N*s.

In the nonpsychosomatic group, the TAT fantasy and Asch conformity measures were positively and significantly correlated ($r_B = +.67$, $p < .05$). This finding cross-validates the results of Kagan and Mussen (1956), and supports the belief that the two measures refer to the same dependency construct. However, these measures were not significantly correlated in the ulcer, nonulcer psychosomatic, or combined psychosomatic groups. The present study allows only for speculation as to why these two measures did not correlate in the psychosomatic groups. Perhaps the relatively high dependency wishes of psychosomatics in connection with the structure of the conformity situation posed a dilemma which they did not resolve consistently from trial to trial. In the nonulcer psychosomatic group, *internalized* conflict arising from dependency-anxiety may have resulted in alternating expressions and inhibitions of the yielding tendency, thereby attenuating the correlation. In the ulcer group, the *external* conflict between trying to please the group by yielding to consensual error and alienating the experimenter by not being as "accurate as possible" could have had the same effect. Further study using more refined measures of conformity is needed to explore these possibilities.

## SUMMARY

This study investigated the hypothesis that Primary Type ulcer patients exhibit strong dependency in both their fantasied and actual social relationships. Subjects were 38 lower-middle- and lower-class male Veterans Administration Hospital patients, subdivided into ulcer, nonulcer psychosomatic, and nonpsychosomatic groups. Dependency fantasy was measured by the number of dependency themes elicited on selected TAT cards, and actual dependency by the number of yields in the Asch group conformity situation. The three groups were compared on both measures.

As expected, the ulcer patients wrote more dependency themes and exhibited greater group conformity than the nonpsychosomatics. The ulcer group was also more conforming than the nonulcer psychosomatics. However, ulcer patients were not significantly more dependent in fantasy than the nonulcer psychosomatics, although the latter group exhibited more dependency fantasy than the nonpsychosomatics.

The pattern of findings suggested that strong dependency needs exist generally in psychosomatic patients, but whereas these needs are often expressed directly by the Primary Type ulcer patient, they tend to be inhibited by nonulcer psychosomatics. Supplementary data on differential dependency-anxiety among the three groups provided tentative support for this interpretation.

The TAT dependency fantasy and Asch conformity measures were correlated in the nonpsychosomatic group, but were not significantly related in the two psychosomatic groups. It was suggested that different forms of dependency conflict in the two psychosomatic groups may have occurred in the conformity situation, thereby attenuating the correlation.

## REFERENCES

ALEXANDER, F. The influence of psychologic factors upon gastro-intestinal disturbances: A symposium. *Psychoanal. Quart.*, 1934, **3**, 501–588.

ALEXANDER, F. *Psychosomatic medicine*. New York: Norton, 1950.

ALEXANDER, F., & FRENCH, T. M. *Studies in psychosomatic medicine*. New York: Ronald, 1948.

ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.

KAGAN, J., & MUSSEN, P. H. Dependency themes on the TAT and group conformity. *J. consult. Psychol.*, 1956, **20**, 29–32.

KAPP, F. T., ROSENBAUM, M., & ROMANO, J. Psychological factors in men with peptic ulcers. *Amer. J. Psychiat.*, 1947, **103**, 700–704.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

MARQUIS, DOROTHY P., SINNETT, E. R., & WINTER, W. D. A psychological study of peptic ulcer patients. *J. clin. Psychol.*, 1952, **8**, 266–272.

MARSHALL, SIMONE. Personality correlates of peptic ulcer patients. *J. consult. Psychol.*, 1960, **24**, 218–223.

MURRAY, H. A. *Thematic Apperception Test manual*. Cambridge, Mass.: Harvard Univer. Press, 1943.

SCODEL, A. Passivity in a class of peptic ulcer patients. *Psychol. Monogr.*, 1953, **67**(10, Whole No. 360).

STREITFIELD, H. S. Specificity of peptic ulcer to intense oral conflict. *Psychosom. Med.*, 1954, **16**, 315–325.

WINTER, W. D. Two personality patterns in peptic ulcer patients. *J. proj. Tech.*, 1955, **19**, 332–344.

(Received December 14, 1960)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 65–68

# UNREALISTIC SELF–ENHANCEMENT IN PARANOID SCHIZOPHRENICS[1]

PHILIP H. HAVENER[2] AND CARROLL E. IZARD
*Vanderbilt University*

In one of the early studies dealing with self-perception, Frenkel-Brunswik (1939) examined the relationship between self-reports and sociometric ratings and observed the tendency of some subjects to utilize a mechanism of defense which she termed distortion into the opposite; e.g., perceiving one's self as being sincere while peers perceived lack of sincerity. In another early study in this area, Sears (1936) described the related defense mechanism of projection as the tendency of subjects to attribute to others certain negative characteristics which they themselves possessed, according to the judgment of their peers. These defense mechanisms of "distortion into the opposite" and "projection" are herein interpreted as over-valuation of self and devaluation of others, defining aspects of the more inclusive defense process of unrealistic self-enhancement.

The self-theory of Rogers (1947, 1950) and Combs and Snygg (1959) leads to the general hypothesis that self-satisfaction and self-acceptance are positively related to adjustment and acceptance of others. Several studies (Rogers & Dymond, 1954) confirmed this hypothesis. However, some researchers failed to find a linear relationship between self-acceptance or self-satisfaction and adjustment (Bennett, 1957; Friedman, 1955). These studies suggested that some types of maladjusted persons, notably paranoid schizophrenics, made unrealistc self-appraisals and were not dissatisfied with themselves to a degree commensurate with the degree of their disturbance. These findings pointed to the tenabilty and explanatory value of the construct of unrealistic self-enhancement.

An underlying assumption of the present study is that unrealistic self-enhancement results from the individual's failure to develop or maintain adequate self-related positive affect and satisfying interpersonal affective ties. Self-related positive affect—positive feelings toward self, self-respect, self-esteem, self-confidence—is seen as a basic need in healthy personality development. As an individual *receives* interpersonal positive affect—positive feelings, respect, esteem—from others, his self-related positive affect increases along with his own tendency to *express* interpersonal positive affect.

It was thought that subjects making extensive use of paranoid defenses would demonstrate a marked tendency to utilize unrealistic self-enhancement—overvaluation of self with concomitant devaluation of others. In the more operational terms of this study, the general hypothesis was that paranoid schizophrenics will show more distortion in the perception of self and others as measured by objective inventories than will persons not using paranoid defenses. The specific hypotheses were:

$H_1$: Paranoid schizophrenic subjects will show a greater difference between expressed acceptance of self and expressed acceptance of others than will normal subjects and nonparanoid schizophrenic subjects.

$H_2$: Paranoid schizophrenic subjects will accept fewer mildly self-derogatory statements about themselves than will normal subjects and nonparanoid schizophrenic subjects.

$H_3$: Paranoid schizophrenic subjects will express a greater amount of self-satisfaction than will normal subjects and nonparanoid schizophrenic subjects.

## METHOD

A group of 20 hospitalized paranoid schizophrenics was selected on the basis of a diagnosis as extreme and chronic in their paranoid behavior and on the basis of a high score on the C (Paranoid Projection), H (Submissiveness vs. Belligerence), and J

[1] This paper is a condensation of part of the senior author's doctoral dissertation, completed at Vanderbilt University under the junior author's supervision.

[2] Now at Orange County Community College, Middletown, New York.

(Self-depreciation vs. Grandiose Exclusiveness) factors of the Scale for Rating Psychiatric Patients (Lorr, 1953). A second group of 20 hospitalized schizophrenics was chosen on the basis of a diagnosis other than paranoid and a low score on the C, H, and J factors of the same scale. A group of 20 attendants at the same hospital formed the second control group. All three groups were equivalent in age, school grade completed, intelligence, and sex.

The Tennessee Department of Mental Health (TDMH) Self-Concept Scale was used to measure distortions in the perception of self (Fitts, 1956, 1959). It consists of 100 statements descriptive of self in five different areas of self-concern, 45 statements being positive, and 45 being negative. Ten additional statements, taken from the MMPI Lie scale, are included to measure a person's honesty in responding. Distortion in the perception of self was measured by two scores from the Self-Concept Scale. One was self-enhancing defensiveness as operationally defined by the *L* score. The second was self-satisfaction as operationally defined by the difference between Row I (abstract description of self) and Row II (satisfaction with self).

The Berger Scale (Berger, 1952), which contains 36 items referring to self and 28 items referring to others, was used to measure distortion in perception of others. Operationally, distortion in the perception of others was defined as a large difference between acceptance of self (AS) and acceptance of others (AO), with the latter as the subtrahend.

## RESULTS

In general, the paranoid schizophrenics demonstrated more distortion in the perception of self and others than did the nonparanoid schizophrenics or the normal group. The data indicated that the paranoid group has an inflated concept of the self and a low concept of others.

To test $H_1$ the mean scores of AS–AO for the first three groups were compared by analysis of variance and *t* tests. The assumptions of normality of distribution and homogeneity of variance were tested by $\chi^2$ for goodness of fit and the $F_{max}$ test (Bliss & Calhoun, 1954), respectively. In both cases the null hypothesis could not be rejected. Using the formula for deriving the reliability of difference scores (Nunnally, 1959), the reliability coefficient for AS–AO was .80. The analysis of variance of the AS–AO scores showed that the group of paranoid schizophrenic subjects had a significantly larger mean score than did the normals and nonparanoid schizophrenic subjects. The *F* was 7.23, with *p* less than .01. The *t* comparing the mean AS–AO score in the paranoid and normal groups was 2.9, which is significant at the .005 level of confidence. The comparable *t* for the paranoid and nonparanoid groups was 3.4, which is significant at the .005 level. These analyses supported $H_1$.

The AS and AO scores of the three groups were compared separately by *t* tests. The mean AS score for paranoids was higher than that of nonparanoid schizophrenics (*t* was 1.97, *p* approached .05) and approximately equal to that of normals. Similarly, on the total net positive (TNP) self-concept score derived from the TDMH scale, paranoids had a mean score approximately equal to that of normals and substantially higher than that of nonparanoids (*t* was 1.82, *p* less than .08). By contrast, the nonparanoids had significantly lower mean scores for AS (*p* less than .01) and TNP (*p* less than .05) than did normals. The mean AO score of the paranoid schizophrenics was significantly lower than that for normals (*t* was 5.46, *p* less than .005); it was lower than that for nonparanoids but not significantly so. Nonparanoids also had a significantly lower mean AO score than normals (*t* was 4.45, *p* less than .005), but the mean difference was not as great as that for paranoids and normals.

To test $H_2$ the mean *L* scores for the three groups were compared by analysis of variance and *t* tests. The assumptions of normality of distribution and homogeneity of variance were tested by $\chi^2$ for goodness of fit and $F_{max}$ test and found tenable. The analysis of variance of the *L* scores showed that paranoid schizophrenics had a significantly smaller mean *L* score than did the normal subjects but not significantly smaller than the nonparanoid schizophrenics. The *F* was 3.86, *p* less than .05. The *t* between the means of the paranoid schizophrenics and the normals was 3.0, *p* less than .01. The *t* between the means of the paranoid schizophrenics and the nonparanoid schizophrenics was 1.1, *p* greater than .05. The mean difference was in the predicted direction. These analyses supported part of $H_2$.

To test $H_3$ the mean self-satisfaction scores (Row I–Row II on the TDMH Self-Concept Scale) for the three groups were compared by analysis of variance and *t* tests. The as-

sumptions of normality of distribution and homogeneity of variance were tested by $\chi^2$ for goodness of fit and the $F_{max}$ and found tenable. The difference score reliability coefficient for Row I–Row II was .78. The analysis of variance of these scores showed that the paranoid schizophrenics had a Row I–Row II score significantly smaller than did the normal subjects but not significantly smaller than the nonparanoid schizophrenics. The $F$ was 3.81, $p$ less than .05. The $t$ for the means of the paranoid schizophrenics and nonparanoid schizophrenics was .86, $p$ greater than .20. The mean difference was in the predicted direction. These analyses partly supported $H_3$.

An additional finding was that the paranoid schizophrenic subjects had a more positive concept of their personal characteristics than of their family relationships, whereas the normal and nonparanoid schizophrenic subjects had a more positive concept of their family relationships than they did of their personal characteristics. This finding was obtained by comparing the three groups on the difference between two subscores of the Self-Concept Scale: Column C (which measures personal characteristics) and Column D (which measures family relationships). The difference-score reliability for Column C–Column D was .67. The data were analyzed by $t$ test. The $t$ for the difference between the means of paranoid schizophrenics and nonparanoid schizophrenics was 2.33, $p$ less than .05; $t$ for paranoid schizophrenics and normals was 2.76, $p$ less than .05. This was a post hoc analysis, since these Self-Concept Scale subscores were not known to the writers at the time the research was planned.

## DISCUSSION

The results showed that the discrepancy between acceptance of self and acceptance of others was significantly greater for paranoid schizophrenics than for normals and nonparanoids. The degree to which paranoids expressed positive self-perceptions (in terms of AS and TNP scores) was approximately equal to that of normals and substantially greater than that for nonparanoid schizophrenics. To what extent this points to veridical self-acceptance in the paranoid schizophrenic is not evident, but it is reasonable to assume that the numerically equal AS and TNP scores of paranoids and normals are not psychologically equivalent. While the paranoids' mean scores on these measures of self-regard were inflated in comparison with those of nonparanoid schizophrenics, their mean scores on the measure of acceptance of others was somewhat lower than that of the nonparanoid schizophrenics. Izard (1959) found that the paranoid not only tends to express unfavorable feeling toward others but actually perceives others (photographs of normals) as hostile, suspicious, and threatening. These results are considered as evidence in support of the basic assumption of the present study that paranoid schizophrenics have inadequate interpersonal affective ties and an overbundance of affect directed toward self and defense of self.

Speculatively, we might say that the paranoid schizophrenic's self-acceptance and self-satisfaction are nonveridical and that, indeed, he cannot experience genuine self-acceptance because he is lacking in self-related positive affect—veridical positive feelings toward self, self-respect, and self-esteem. His deficiency of self-related positive affect developed as a function of insufficient interpersonal rewards, which in turn is related to his tendency to devalue other people and to perceive them as hostile and threatening. Thus, his unrealistic self-enhancement—overvaluing of self and devaluing of others—can be seen as a defense against complete loss of self-related positive affect and of interpersonal positive affect.

## SUMMARY

The basic assumption underlying this study was that adequate self-related and interpersonal positive affect are necessary for effective adjustment. The general hypothesis was that paranoid schizophrenics would show greater distortion in the perception of self and others than nonparanoid schizophrenics or normals. The specific hypotheses are listed below:

$H_1$: Paranoid schizophrenic subjects will show a greater difference between expressed acceptance of self and expressed acceptance of others than will normal subjects and nonparanoid schizophrenic subjects.

$H_2$: Paranoid schizophrenic subjects will accept fewer mildly self-derogatory statements about themselves than will normal subjects and nonparanoid schizophrenic subjects.

$H_3$: Paranoid schizophrenic subjects will express a greater amount of self-satisfaction than will normal subjects and nonparanoid schizophrenic subjects.

Twenty paranoid schizophrenics, 20 nonparanoid schizophrenics, and 20 normals were given the Berger Scale and the Tennessee Department of Mental Health Self-Concept Scale. These inventories furnished indices of acceptance of self (AS), acceptance of others (AO), self-satisfaction, and an *L* score. The three hypotheses were tested by comparing the three groups on AS–AO, self-satisfaction, and *L* scores by means of analysis of variance and *t* tests. The results confirmed $H_1$ and part of $H_2$ and $H_3$. An additional analysis of two subscores of the Self-Concept Scale further substantiated the general hypothesis by showing that paranoid schizophrenics had higher opinions of their personal characteristics and lower opinions of their family relationships than did normals and nonparanoid schizophrenics.

The results showed that the paranoids tended to overrate self as compared with nonparanoid schizophrenics. It was reasoned that this was evidence of unrealistic self-enhancement, a defense against complete loss of genuinely positive self-related affect and of satisfying interpersonal affective ties.

## REFERENCES

BENNETT, J., JR. Modification of the self-concept in electroshock therapy. *Dissertation Abstr.*, 1957, **17**, 146.

BERGER, E. M. The relation between expressed acceptance of self and expressed acceptance of others. *J. abnorm. soc. Psychol.*, 1952, **47**, 778–782.

BLISS, C. I., & CALHOUN, D. W. *An outline of biometry*. New Haven: Yale Cooperative Corp., 1954.

COMBS, A. W., & SNYGG, D. *Individual behavior*. New York: Harper, 1959.

FITTS, W. H. The Tennessee Department of Mental Health Self-Concept Scale: Preliminary manual. Nashville, Tenn.: Tennessee Department of Mental Health, 1956. (Mimeo.)

FITTS, W. H. The self-concept and mental health: A research report on the Tennessee Department of Mental Health Self-Concept Scale. Nashville, Tenn.: Tennessee Department of Mental Health, 1959.

FRENKEL-BRUNSWIK, ELSE. Mechanisms of self-deception. *J. soc. Psychol.*, 1939, **10**, 409–420.

FRIEDMAN, IRA. Phenomenal, ideal and projected conceptions of self. *J. abnorm. soc. Psychol.*, 1955, **51**, 611–615.

IZARD, C. E. Paranoid schizophrenic and normal subjects' perception of photographs of human faces. *J. consult. Psychol.*, 1959, **23**, 119–124.

LORR, M. Multidimensional scale for rating psychiatric patients, hospital form. *V. A. Tech. Bull.*, 1953, No. 10-507.

NUNNALLY, J. C., JR. *Tests and measurements*. New York: McGraw-Hill, 1959.

ROGERS, C. R. Some observations on the organization of personality. *Amer. Psychologist*, 1947, **2**, 358–368.

ROGERS, C. R. The significance of self-regarding attitudes and perception. In M. L. Reymert (Ed.), *Feelings and emotions: The Mooseheart symposium*. New York: McGraw-Hill, 1950. Pp. 374–382.

ROGERS, C. R., & DYMOND, ROSALIND F. *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954.

SEARS, R. R. Experimental studies in projection. *J. soc. Psychol.*, 1936, **7**, 151–163.

(Received December 19, 1960)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 69–73

# THE MEASUREMENT OF CHARACTERISTIC PATTERNS OF WARD BEHAVIOR IN CHRONIC SCHIZOPHRENICS

MARVIN HUNTER [1]

*Veterans Administration Hospital, Brooklyn, New York*

CARMI SCHOOLER [2]

*National Institute of Mental Health*

AND HERBERT E. SPOHN

*Veterans Administration Hospital, Montrose, New York*

At the present time, there is relatively little systematic fully objective knowledge about the overt behavior of chronic schizophrenic individuals hospitalized in large neuropsychiatric institutions.

Existing techniques for the measurement of overt patient behavior are typically rating scales in which various conceptions of psychopathology have been dimensionalized. The most prominent and probably most widely used of these is Lorr's Multidimensional Psychiatric Rating Scale (Lorr, Jenkins, & O'Connor, 1955). There are also measurement techniques developed for special purposes to record particular aspects of patient behavior (Laffal, Sarason, Ameen, & Stern, 1957; Pace, 1957). Lacking are measuring instruments free of any assumptions about how psychopathology is organized in neuropsychiatric patients and which are sufficiently comprehensive to yield a full, objective picture of the patients' adjustment in the hospital environment. In other words, what is needed to provide a firm basis for projecting hypotheses about the nature and structure of the behavior of neuropsychiatric patients are measuring instruments which tell us more about what the patient actually does or does not do, rather than instruments which tell us where his behavior places him on a continuum of psychopathology.

The lack of such objective, comprehensive methods for assessing patient behavior becomes especially apparent when one confronts the task of measuring and evaluating the effects of a socio-environmental treatment program. This lack provided the impetus for the development of the observational instrument, the Location-Activity-Inventory (LAI),[3] to be described here. This article is concerned with a description of the development of this instrument, its major relevant characteristics and use, its reliability and finally the kinds of patient "images" that analysis of the LAI data is capable of yielding.

The LAI was designed to provide knowledge about the behavior of chronic schizophrenics as relevant to the environment in which they characteristically function—the ward of a large neuropsychiatric hospital. Specifically, the instrument was developed to measure the behavioral characteristics of a group of 100 male patients on two closed wards of a neuropsychiatric hospital and to study their responses to a projected socio-environmental treatment program. The major objective in its development was to obtain, through a time sampling method, a reliable and comprehensive description of the overt ward behavior of chronic schizophrenic patients.

Extensive free hand observations had dis-

[1] Hunter carried out this study while at the Veterans Administration Hospital, Montrose, New York.

[2] Schooler carried out this study while at the Veterans Administration Hospital, Montrose, New York.

[3] Copies of the manual detailing the method of observation and the operational definitions of observational categories have been deposited with the American Documentation Institute. Order Document No. 6940 from ADI Auxiliary Publication Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

closed four major dimensions of patient ward behavior as relevant: (*a*) geographical location in the ward setting; (*b*) position in relation to the wall; (*c*) posture or body disposition; and (*d*) activity or response to ward stimuli.

These particular dimensions were at length chosen to represent patient behavior because they seemed to discriminate between patients and because they appeared to reflect the terms in which patients make their characteristic adjustment, both pathological and nonpathological to the ward environment. Furthermore, it was felt that they would yield behavioral profiles permitting systematic inferences regarding the dynamics of behavior and pathology.

The specific categories used to describe behavior characterizing the four dimensions were selected on the basis of three criteria: first that the need for subjective judgment on the part of an observer be minimal; second, that the categories be exhaustively descriptive; and, third, that the categories be as specific as possible without ceasing to represent meaningful units of behavior.

As regards the Location dimension the total ward space was subdivided into 14 functional geographical areas. The presence of a patient in one of these areas when observed thus constituted his location response. The Position dimension was characterized in terms of six response categories: "Against the Wall," "Facing the Wall," "Against Corner," "Facing Corner," "Looking out Window," and "None of the Above." The Posture categories consisted of "Sitting," "Standing," "Lying," "Walking," "Sitting on the Floor," and "Lying on Floor." Finally, the categories descriptive of the Activity dimension; i.e., patient responses to the varied stimuli of the ward environment were divided into several broad classes. These are (*a*) social activities (verbal-social, gestural-social, laughing-social, physical contact, and social games); (*b*) parasocial activities (reading, writing, nonsocial games, TV viewing, verbal, laughing, and gestural parasocial); (*c*) functional, nonsocial activities (ward service, personal care, excretory, and functional object manipulation); (*d*) nonfunctional activities (self-manipulation, active movement, nonfunctional object manipulation, verbal, gestural, and laughing nonsocial); (*e*) no behavior ("null" behavior and sleep).

While these categories may seem to be face-valid and obvious, precise definitions of each were required in order to obtain interobserver agreement. For example, a patient was judged to be "against the wall" if his body or the back of the chair in which he was seated was within 2 feet of the wall. The care that had to be exercised in this regard is perhaps even better illustrated by the definition of "television viewing": "A subject will be judged as watching television in the presence of either a specific posture response or an observable empathic or reactive response to the television set. The posture response refers to a position of the body which requires some special strain to orient the head toward the television set."

## PROCEDURE

The LAI is used by a single observer. A trained observer requires about 15 minutes to record the behavior of 50 patients on a ward. He moves through the ward along a predetermined route. In each of the 14 areas of the ward, he notes the number and identity of the patients present as he enters. He then observes each patient for a maximum of 10 seconds and records behavior using a single category in each of the four major dimensions. The categories are regarded as mutually exclusive. When a patient is engaged in more than one activity, the observer makes a determination of which is the most salient. After recording the behavior of patients in a given area, the observer moves on to the next area and repeats the procedure until all areas have been covered. In general, the observation procedure is roughly analogous to taking a panoramic photograph of all the patients on the ward, with a 10-second exposure and using a film sensitized to pick up specified dimensions of behavior.

Observers did not make data-yielding observations with the LAI until they had made practice observations on the wards for some time and until their presence on the ward had become commonplace from the patient's point of view. The time, the order in which observers moved through the geographical areas of the ward, and the observation posts adopted by the observers were varied for each occasion of observation to minimize the predictability of the observer's appearance on the ward or his pattern of movements.

## RESULTS

The data reported in this article represent 6 months of observation by four observers on two wards. These observations were made

TABLE 1

PERCENTAGES OF TOTAL FREQUENCIES OF OBSERVATIONS IN THE VARIOUS CATEGORIES

| Position | | Posture | | Activities | |
|---|---|---|---|---|---|
| Against wall | 46.0 | Sitting | 61.1 | Null behavior | 44.0 |
| Against corner | 4.6 | Standing | 17.2 | Sleeping | 10.0 |
| Facing wall | .8 | Lying | 6.7 | Social (Interaction) | 2.7 |
| Facing corner | .3 | Walking | 13.1 | Parasocial | 15.4 |
| Looking out window | 1.4 | Sitting on floor | 1.1 | Functional Nonsocial | 9.2 |
| None of the above | 45.4 | Lying on floor | .3 | Nonfunctional (Pathological) | 18.7 |
| No recording | 1.5 | No recording | .7 | | |
| Total | 100.0 | Total | 100.2 | Total | 100.0 |

during the base-line phase of an experimental ward project while the wards were functioning under a characteristic custodial regimen. During this period each waking hour of the patient's day (total = 12) and each day of the week were equally sampled. Observations were scheduled in such a way as to avoid systematic time of day or day of week bias. An average of 39 patients present on each ward was observed on 240 occasions yielding a total of approximately 9,000 patient occasions of observation or an approximate average of 90 observations for each of the 100 patients observed.

The reliability of the instrument was examined from two points of view: degree of agreement among observers and split-half reliability of categories based on correlations between odd and even occasions of observation. Observer reliability was established by randomly pairing observers and having them make joint but independent observations. Contingency coefficients for each pair were calculated for the position, posture, and activity categories. The contingency coefficients for position averaged .84, for posture .86, and .78 for the activity dimension. The split-half reliability coefficients calculated for individual categories averaged .91 across the categories. (These correlations are Pearson $r$'s corrected by the Spearman-Brown prophecy formula for the full length of the scale). These analyses indicate a reasonable degree of observer agreement in making behavioral judgments and a high degree of stability of patient behavior on a group basis.

The behavioral profile characterizing the chronic schizophrenic patients observed with the LAI is shown in Table 1. The percentages shown in the various categories represent the proportions of total frequencies cumulated by all patients in the position, posture, and activity dimension.[4]

The most striking aspect of the above table is found in an examination of the activity categories, in which one finds that patients spend 44% of their time in null behavior, (total immobility) and 11% in sleeping. This means that during the waking day on the ward, patients spend more than half of their time in the absence of any kind of overt activity neither talking nor reading, nor even scratching themselves. In contrast to the large

[4] It should be noted that the proportions shown in the table represent cumulated frequencies for the total group of patients and do not reflect the contributions of individuals to the total percentage or interindividual variability.

The group of 100 male patients characterized by these data were all diagnosed as schizophrenic (29 paranoid, 37 hebephrenic, 14 catatonic, 10 simple, 10 undifferentiated or mixed). They had been hospitalized for an average of 8.9 years and had an average age of 39.8 years. No patients with known organic brain damage were included.

TABLE 2

PERCENTAGES OF SOCIALIZERS AND NONSOCIALIZERS DISPLAYING NULL, PARASOCIAL, AND PATHOLOGICAL BEHAVIOR

| Behavioral category | Socializers | Nonsocializers | $t$ |
|---|---|---|---|
| Null | 35.72 | 49.52 | 3.6* |
| Parasocial | 19.50 | 12.98 | 2.1* |
| Pathological (nonfunctional) | 14.72 | 19.99 | 1.7 |

* Significant at <.05 level of confidence.

amount of time doing nothing, the patients spend only 2.7% of their time in social activities. Furthermore, this figure must be interpreted with care, for the median frequency for all social behavior is zero.

While not all possible interrelationships among categories and between categories and other independent measures of behavior have been examined, the following are presented to exemplify the psychologically consistent picture of patient behavior which emerges upon internal examination and the rational relationships of behavior as observed on the LAI with other levels of behavioral measurement. Thus for example, patients who are characterized by a relatively high rate of social interaction differ from nonsocializers in other characteristics as well.

As can be seen from Table 2, those who were observed in at least one social interaction differ significantly from those who were never observed in such an interaction in several ways.[5] Those who socialized exhibited more parasocial and less null behavior. Furthermore, there is some evidence that socializers also exhibited less nonfunctional pathological appearing behavior. Although the difference between the two groups in this latter behavior shown in the table did not achieve significance, a further analysis in which the group of socializers were divided by a median break, revealed that the high socializers were significantly different from those who did not socialize. (High socializers, mean 12.88; nonsocializers, mean 19.99, $t = 1.98$, $p < .05$).

As regards relationships between measurements of overt behavior via the LAI and other levels of behavioral measurement—such as "intrapsychic" factors, a comparison with Rorschach data was made using the Becker (1956) Developmental Level Score (DL). This score is based on a one-response-per-card Rorschach in which a high score is interpreted as representing a high level of perceptual organization and by inference a high level of personality integration. The $N$ here is 75. A number of patients proved untestable.

Significant correlations ($p < .05$) were found between DL score and social behavior (.273); between DL and parasocial behavior (.340); between DL and nonfunctional behavior (− .313); between DL and null behavior (− .292). These correlations seem to confirm the expectation that the segments of behavior measured by the LAI are not unique and isolated phenomena but bear significant relationships to other behavioral modalities.

One specialized analysis undertaken with LAI was designed to test the popular belief that the phases of the moon affect the behavior of the mentally ill patients. While this may appear frivolous, it should be pointed out that the lunacy hypothesis is widely believed by personnel and that systematic changes in patient behavior related to phases of the moon might reflect personnel expectations. For purposes of this analysis observational data were divided into four quarters corresponding in time to occasions of the full moon, the last quarter, the new moon, and the first quarter.

A comparison of relevant indices such as nonsocial verbal and other tension reflecting pathological behaviors, and of null behavior reveals no significant differences in frequency of these behaviors associated with the various phases of the moon. While not a conclusive test of the "lunacy" hypothesis, the analysis at least reflects a remarkable degree of consistency in the behavior of schizophrenic patients.

Data on daily temperature, humidity and prevailing weather conditions were also gathered and related to several relevant indices. Again no significant differences systematically related to variations in these conditions were obtained. Two points can be made in this connection. The observational data were gathered during a period from October 1957 to March 1958. The ranges of weather variations during this period is considerably narrower than the total range of climatic variation throughout all seasons in the region where the observations were made. Had greater extremes of weather conditions, etc. been included differences in behaviors associated with them might have emerged. It is conceivable, on the other hand, that chronic schizophrenics are relatively oblivious to external weather conditions. In

[5] The entries in this table are the average percentages based on individual percentages in the several groups.

any event the data at hand do not permit a conclusive test of this assertion.

Some as yet unanswerable questions regarding the usefulness of this instrument in measuring *behavioral change* may be raised. It is not known how sensitive it is to change, i.e., how acute the instrument is. The behavioral segments represented by the descriptive categories may or may not be too gross to pick up subtle changes in functioning. Consisting as it does, however, of relatively specific descriptive categories, change patterns that do emerge on the LAI may be examined in some detail. At the same time, the concreteness of the behavioral categories tends to safeguard against the possibility of observer stereotypes as well as against the influence of observer bias when behavioral change is anticipated.

Finally the LAI seems well suited to use and application in a variety of neuropsychiatric settings with only minor modifications of categories to accommodate local conditions of geography and stimulus environment. Wider application at this juncture would provide a means, among other things of exploring systematically the effects of different hospital environments upon the behavior of schizophrenic patients. Similarly, the instrument might prove of considerable interest in making comparisons between a variety of populations. Moreover, the instrument should prove useful in studies where a detailed, objective description of patient behavior in a natural environmental setting is desired.

## SUMMARY

The Location-Activity-Inventory was developed to measure objectively the ward adjustment of neuropsychiatric patients. It is a nominal scale, which on the basis of systematic observations provides behavioral profiles of patients in the areas of patterns of patient location on the ward, characteristic orientation to major architectural features, posture, and patterns of activities or response to the social and nonsocial stimuli on the ward. Both intraobserver reliability coefficients and split-half reliability coefficients for individual descriptive categories are sufficiently high to assume a reasonable degree of stability of measurement.

About 9000 patient observations of 100 deteriorated chronic schizophrenic patients obtained with this scale over a period of 6 months reveal a remarkable dearth of social interaction and a preponderance of immobility in such patients.

Psychologically meaningful covariances among descriptive categories were observed as well as correlations of ward behavior with clinical data suggesting that the instrument measures more than isolated, unique segments of behavior. Applications for use of the instrument in evaluative studies and in a variety of settings for comparative purposes are suggested.

## REFERENCES

BECKER, W. A genetic approach to the interpretation and evaluation of the process-reactive distinction in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **53**, 229–236.

LAFFAL, J., SARASON, I. G., AMEEN, L., & STERN, A. Individuals in groups. *Int. J. soc. Psychiat.*, 1957, **4**, 254–262.

LORR, M., JENKINS, R. L., & O'CONNOR, J. P. Factors descriptive of psychopathology and behavior of hospitalized psychotics. *J. abnorm. soc. Psychol.*, 1955, **50**, 78–86.

PACE, R. E. Situational therapy. *J. Pers.*, 1957, **25**, 578–588.

(Received December 19, 1960)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 74–78

# A VALIDATION STUDY OF THE PROCESS SCALE[1]

T. M. TOMLINSON AND J. T. HART JR.
*University of Wisconsin*

In a 1958 article Rogers presented a theoretical and descriptive formulation of some of the process changes in personality and personal expression which occur during successful psychotherapy. Since the publication of that article a number of investigations have been conducted to empirically elucidate and operationalize Rogers' formulation (Hart, 1961; Rogers, 1959; Tomlinson, 1959; van der Veen, 1961; Walker, Rablen, & Rogers, 1960).

The first step in this line of research was the development by Rogers and Rablen (1958) of a scale to quantify the process conception of personality.[2] As described by Rogers (1959), the rating scale portrays aspects of client functioning in therapy along a general continuum extending from "rigidity and fixity of psychological functioning on the one hand to psychological flow and changingness on the other."

For a detailed description of the Process Scale the reader should consult the article by Walker, Rablen, and Rogers (1960). The measure can be briefly characterized, for the purposes of this paper, as a seven-stage standard scale for the rating of typescript and/or tape excerpts from therapy interviews. The scale as a whole is divided into seven vertical subcategories or strands—Feelings and Personal Meanings, Experiencing, Incongruence, Communication of Self, Construing of Experience, Relationship to Problems, and Manner of Relating. When rating, the raters take each strand or subscale into account and arrive at a global rating of the client's process level which may range from 1.0 to 7.0.

[1] This investigation was supported (in part) by research fellowship awards from the United States Office of Vocational Rehabilitation, Washington, D. C. and by a research grant from the National Institute of Mental Health (as part of a larger project), Grant No. M-3496.

[2] Mimeographed copies of the latest revision of the Process Scale may be obtained by sending a request to the authors at the Psychiatric Institute, University of Wisconsin, Madison, Wisconsin.

This paper reports the latest investigation in the series of studies which have been carried out to determine the reliability and validity of the Process Scale.

## HYPOTHESES

For the Process Scale to be a useful instrument for research and assessment it must, in the hands of competent judges, yield the following kinds of information. First, a scale should be able to discriminate, by means of obtained ratings, a more successful case from a less successful case.

Secondly, and possibly most important, there should be a difference between the obtained ratings for early as opposed to late interviews. These differences should be in the direction of later interviews yielding a higher process rating than earlier interviews. Further, the differences should be more pronounced for interviews from a successful case than for interviews from a less successful case.

Carrying the above one step further, if the scale is sensitive to the hypothesized changes which occur as a function of psychotherapy, and if psychotherapy is a reasonably continuous phenomenon, then there should be differences, though small, between first and second halves of an interview. Again the differences should be greater for a more successful case than for a less successful case.

The above can be summarized in the form of hypotheses as follows:

1. Process ratings for more successful cases, hereafter abbreviated m.s., as rated by multiple criteria of outcome, should be significantly higher than process ratings for less successful cases, hereafter abbreviated l.s.

2. Process ratings should be significantly higher for later interviews than for early interviews.

Corollary 2a: Process ratings should show more pronounced differences between early and late interviews of a m.s. case than between early-late interviews for a l.s. case.

3. Process scores will be higher for the second half than for the first half of a given interview, although this intrainterview difference will be smaller than the interinterview difference.

Corollary 3a: Process ratings should show more pronounced differences between first and second halves of interviews in a m.s. case than between halves of a l.s. case.

## METHOD

### *Cases and Rating Segments*

A sample of 10 cases, considered representative of the University of Chicago Counseling Center clients, was selected. For each of the 10 cases an early interview and a late interview tape recording was obtained; the second interview from the beginning and the next to the last interview. Nine 2-minute samples of therapy were systematically selected from each of the 20 tapes. The sampling units were arbitrarily chosen at every 3-minute interval within the interview, e.g., 3′interval-2′unit-3′interval-2′unit-3′interval-2′unit, and so on until nine units had been selected from each tape. After selection these 180 units were coded to disguise their origin from the raters, and in addition, the 180 units were randomly assigned to 18 rating groups with 10 rating units per group. For every 2-minute unit a tape recording and a typescript of the tape recording were given to the judges for ratings.

### *Raters*

Two raters were used both of whom had participated in previous studies of the Process Scale. These two experienced raters rated all 180 units—assigning one rating to each 2-minute sample. The rating period extended over 18 days since each rater rated 10 of the 180 units per day. The raters made the ratings at separate times and were not allowed to discuss their ratings with one another—or with others. Both raters had previous experience in psychotherapy and were considered sensitive to intherapy events.

### *Procedure*

The raters used both tape recordings and typescripts to make their ratings of the 2-minute units. Typescripts were necessary because of the poor fidelity of some of the tape units. Tape recordings were used in addition to the typescripts to provide such cues as tone quality, rate of speech, emphasis, etc.

Of the 10 cases selected, 5 were considered more successful cases and 5 were considered less successful cases on the basis of a multicriteria success score. The criteria included therapist ratings of outcome, patient ratings of outcome, and a self-concept *Q* sort. The m.s. and l.s. cases were randomly assigned to rating groups and all cues in the tape segments concerning the number of the interview or the satisfaction of the client with the therapy were eliminated.

The judges were instructed to rate each unit as a separate entity and to try not to relate an earlier rating to a later rating. The judges reported no influence in their rating due to hearing previous segments from the same case, however this was impossible to assess objectively within the present design.

TABLE 1

ANALYSIS OF VARIANCE OF PROCESS RATINGS

| Source | *df* | *SS* | *MS* | *F* |
|---|---|---|---|---|
| Success (S) | 1 | 11.55 | 11.55 | 40.10** |
| Judges (J) | 1 | 7.08 | 7.08 | 26.04** |
| Interviews (I) | 1 | 1.57 | 1.57 | 6.00* |
| Halves (H) | 1 | .14 | .14 | .53 |
| S × J | 1 | .11 | .11 | .42 |
| S × I | 1 | .80 | .80 | 3.05 |
| S × H | 1 | .12 | .12 | .46 |
| I × J | 1 | .228 | .22 | .84 |
| H × J | 1 | .05 | .05 | .19 |
| H × I | 1 | .01 | .01 | .04 |
| H × S × J | 1 | 1.44 | 1.44 | 4.75* |
| I × S × J | 1 | .28 | .28 | 1.07 |
| H × I × S | 1 | .97 | .97 | 3.70 |
| H × I × J | 1 | .30 | .30 | 1.15 |
| H × I × J × S | 1 | 1.52 | 1.52 | 5.81* |
| SS′/Success | 8 | 2.30 | .288 | 1.10 |
| Residual | 56 | 14.70 | .262 | |

*$p < .05$.
**$p < .01$.

## RESULTS AND DISCUSSION

Using the Pearson method of correlation, a coefficient of .65 was obtained between the two judges. Considering the complexity of the judgmental task this figure was accepted as adequate. However, it is not a high reliability figure and the results presented hereafter should be interpreted accordingly.

Before presenting the main body of results and discussion it should be pointed out that the variables "Interviews" and "Halves" were not independent, and the reader should interpret the findings within this limitation.

The results then, which are summarized in Table 1 are as follows:

1. Success: The variable success is significant beyond the .01 level in the direction of the m.s. cases having a higher process rating than the l.s. cases and thus supports Hypothesis 1. This highly significant differ-

ence between success groups supports previous results by Walker et al. (1960) and Tomlinson (1959), and has proven to be the most stable finding in this series of studies. This means that the Process Scale is able to discriminate, in a highly reliable way, cases rated m.s. from those rated l.s.

2. Interviews: The scale was able to detect differences between early and late interviews at beyond the .05 level thus supporting Hypothesis 2. Process ratings were, on the average, higher for the later interview than for the early interview.

3. Interview × Success: Corollary 2a is not statistically supported ($p > .05$), however, the result is in the predicted direction. Tentatively process ratings increase in magnitude as a function of success and time of occurrence, e.g., the late, m.s. group received the highest mean rating. Table 2 illustrates this statement more clearly.

4. Halves: There is no statistical support ($p > .05$) for the contention that process scores will be higher for second than for first half of an interview. However, Table 2 shows that all means are in the predicted direction with one exception, and it can hardly be called a major reversal of prediction.

5. Halves × Success: Corollary 3a is not statistically supported ($p > .05$), but the trend seems clearly supportive of the prediction. Table 2 shows that the differences between m.s. halves are relatively clear as compared to the differences between l.s. halves of interviews.

6. Judges: The difference between judges is highly significant, ($p < .01$). When considered in light of the interjudge reliability, this result indicates that although judges appear to be able to rate up and down the scale in a fairly reliable way, they are subject to error which makes them consistently different in the rating they assign to a given segment. It is believed that this differential bias is the reason for the significant second and third order interactions. Plotting the mean process ratings for both judges indicates that there is a generally increasing linear trend in process ratings as a function of time and success, but the magnitude of the interaction depends on which judge makes the rating.

It is not surprising that the difference between halves of interviews is not significant. Successful psychotherapy is at least in part a function of the number of interviews

TABLE 2

MEAN PROCESS RATINGS, NUMBER OF INTERVIEWS, AND COUNSELOR RATINGS OF OUTCOME FOR EACH OF TEN CASES AND TWO JUDGES

| Case | No. of interviews | Counselor rating | Judge 1 | | | | Judge 2 | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Early | | Late | | Early | | Late | |
| | | | 1/2 | 2/2 | 1/2 | 2/2 | 1/2 | 2/2 | 1/2 | 2/2 |
| 1 | 45 | 7 | 3.6 | 4.1 | 4.0 | 4.4 | 2.5 | 2.6 | 2.4 | 2.6 |
| 2 | 46 | 8 | 2.8 | 3.2 | 2.9 | 3.1 | 2.6 | 2.7 | 2.4 | 2.6 |
| 3 | 21 | 8 | 3.5 | 3.8 | 5.2 | 5.0 | 3.9 | 4.1 | 4.6 | 4.8 |
| 4 | 25 | 8 | 3.4 | 3.1 | 4.4 | 4.8 | 2.7 | 2.4 | 3.4 | 3.5 |
| 5 | 14 | 6 | 3.7 | 3.8 | 3.8 | 4.0 | 3.5 | 3.2 | 3.1 | 3.7 |
| Mean | | | 3.4 | 3.6 | 4.1 | 4.3 | 3.0 | 3.0 | 3.2 | 3.4 |
| 6 | 28 | 3 | 3.2 | 3.3 | 3.8 | 3.9 | 2.9 | 2.6 | 2.8 | 2.9 |
| 7 | 27 | 2 | 2.7 | 2.8 | 2.4 | 2.6 | 2.2 | 2.2 | 2.4 | 2.1 |
| 8 | 27 | 5 | 3.4 | 3.2 | 3.4 | 3.5 | 2.6 | 2.4 | 2.5 | 2.6 |
| 9 | 26 | 3 | 2.7 | 2.8 | 2.7 | 2.4 | 2.2 | 2.4 | 2.5 | 2.5 |
| 10 | 38 | 5 | 2.5 | 2.8 | 2.8 | 3.0 | 2.4 | 2.6 | 2.5 | 2.2 |
| Mean | | | 2.9 | 3.0 | 3.0 | 3.1 | 2.5 | 2.4 | 2.5 | 2.5 |

(Standal & van der Veen, 1957) and it would be unusual if within any single interview there was a sudden, perceptible personality change. For this reason no significant difference in mean scores for halves was expected. However, if the process of psychotherapy is a fairly continuous phenomenon, it would be reasonable to expect consistent differences between halves, in the direction of the second half being higher in process rating than the first half. To test this notion a nonparametric statistical procedure (sign test) was used. The differences, holding Success and Judges constant, are in the predicted direction 13 of 19 times ($p = .084$) for the early interviews and 15 of 19 times ($p = .010$) for the late interviews. Holding Judges and Time constant, 16 of the 20 ($p = .006$) interviews are in the hypothesized direction for the more successful cases while 12 of 18 ($p = .119$) are in the positive direction for the less successful cases. Thus it seems that movement within the interview is both a function of time and success. From these results it would seem that the process of personality change within an interview is directly analogous to that which occurs between interviews, i.e., during the course of therapy. The only difference is in the magnitude of the change. Further, the results suggest that the change is a cumulative one whose direction and possibly magnitude varies postively as a function of success and time.

If personality change is a function of psychotherapy and if the Process Scale is sensitive to the behavioral differences that occur concomitantly with this change, then one would expect these changes to be most noticeably reflected in the differences between early and late successful interviews. Table 1 shows this to be the case statistically, but the magnitude (See Table 2) of the differences is hardly of practical value, being on the average only .7 of a stage for Judge 1 and .3 of a stage for Judge 2. Yet Walker, et al., were able to detect mean change scores of 1.93 stages in the marked success group ($N = 3$) using the same scale. However, Walker used cases chosen for the purpose of representing end-points of a success-failure continuum. The cases used in this study cover the entire continuum of success-failure. As such it is reasonable to expect overlap between m.s. and l.s. cases in the middle of the continuum. Further, in analyzing the results in terms of means the differences within the individual success and/or failure cases are washed out. Difference scores were computed for each case between early and late interview ratings and the results suggest that the Process Scale detects differences between early and late interviews at least in part as a function of a success-failure continuum. For Judge 1 the differences between early and late, m.s. interviews range from 1.5 to .0 stages, while for Judge 2 the range is .8 to −.1 stages. For the l.s. group the range for Judge 1 is .6 to −.3 stages and for Judge 2, .2 to −.1 stages. Thus it appears that the Process Scale is sensitive to change as a function of success-failure.

One final comment: This study is the latest in a series which have attempted to validate the Process Scale. Out of these studies one clear finding has always resulted, namely that the m.s. case almost invariably starts therapy at a higher stage of process than the l.s. case. In this study, using the Mann-Whitney $U$ test to test the significance of the difference between early, m.s. and early, l.s. cases, the differences were significant in the expected direction at about the .02 level ($U = 20.5$; $N_1 = 10$, $N_2 = 10$). Here as in the previous studies, the m.s. clients seem to enter therapy at a higher stage of process than the l.s. ones. This suggests at least two things: ($a$) there may be some personality characteristics which each client must possess in some minimal degree to profit from psychotherapy, and ($b$) those clients below this minimum will not progress in psychotherapy. This conjecture is similar to and supports Kirtner's (1955) finding that the successful client is characterized by different personality attributes than the client who fails to profit from psychotherapy. If this approach continues to be fruitful then perhaps more light can be shed on the nature of and the reasons for the failure case.

## SUMMARY

The steps which have been taken to objectify and validate Rogers' conception of the process of psychotherapy have been sum-

marized. The study currently reported, and the previous studies justify the following conclusions in regard to the Process Scale and its operational meaning.

1. Interjudge agreement is always adequate and significantly better than chance, as estimated by parametric and nonparametric statistics. Interjudge reliability appears to have a minimum of about .60.

2. It distinguishes between more successful and less successful cases at a high level of statistical significance.

3. It indicates that more successful cases begin as well as end at a significantly higher level of process.

4. There is evidence that there is greater movement (process change) on the Process Scale during the period of therapy in more successful than in less successful cases. The difference is only modestly significant and its utility is yet to be determined.

5. There is a tendency for the second half of each interview to be rated higher on the Process Scale than the first half. This tendency toward intrainterview change is definitely significant in the more successful group of cases.

## REFERENCES

HART, J. T. A reliability study of the Process Scale. Unpublished MS thesis, University of Wisconsin, 1961.

KIRTNER, W. H. Success and failure in client-centered therapy as a function of personality variables. Unpublished MA thesis, University of Chicago, 1955.

ROGERS, C. R. A process conception of psychotherapy. *Amer. Psychologist*, 1958, **13**, 142–149.

ROGERS, C. R. A tentative scale for the measurement of process in psychotherapy. In E. A. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy*. Washington, D. C.: American Psychological Association, 1959. Pp. 96–107.

ROGERS, C. R., & RABLEN, R. A. A scale of process in psychotherapy. Unpublished manual, University of Wisconsin, 1958. (Mimeo.)

STANDAL, S. W., & VAN DER VEEN, F. Length of therapy in relation to counselor estimates of personal integration and other case variables. *J. consult. Psychol.*, 1957, **21**, 1–9.

TOMLINSON, T. M. A validation study of a scale for the measurement of the process of personality change in psychotherapy. Unpublished MS thesis, University of Wisconsin, 1959.

VAN DER VEEN, F. Strand scale analysis of the psychotherapy Process Scale. *Psychiat. Inst. Bull.*, 1961, **1**, No. 1.

WALKER, A., RABLEN, R. A., & ROGERS, C. R. Development of a scale to measure process change in psychotherapy. *J. clin. Psychol.*, 1960, **1**, 79–85.

(Received December 22, 1960)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 79–83

# NEED FOR SOCIAL APPROVAL AND THE OPERANT CONDITIONING OF MEANINGFUL VERBAL BEHAVIOR

DAVID MARLOWE[1]

*Dartmouth College*

The verbal conditioning paradigm devised by Greenspoon (1955) along the lines of Skinnerian operant conditioning procedures has stimulated a considerable interest in factors related to the control of human verbal behavior. Recent studies have clearly demonstrated that a variety of verbal behaviors can be conditioned by operant techniques, that various kinds of "social reinforcers" are effective, and that conditioning of verbal behavior occurs in the absence of any awareness on the subjects' part as to the "response-reinforcement" relationship (Krasner, 1958; Salzinger, 1959). While it has been almost axiomatic to state that there are marked individual differences in susceptibility to verbal conditioning, there have been conspicuously few reported studies bearing on possible personality or motivational correlates of verbal conditioning. Studies of the relationship of anxiety level to conditionability have yielded equivocal results (Rogers, 1960; Taffel, 1955). The present study was undertaken in order to examine the influence of "need for social approval" on the conditioning of verbal behavior in a meaningful social situation.

The construct of need for social approval and the scale developed to measure this variable were originally formulated within the psychometric framework of a socially desirable response set (Crowne & Marlowe, 1960; Marlowe & Crowne, 1961) with the intent to account for individual differences in the strength of the tendency to respond in a socially desirable manner, i.e., in accord with situational and cultural demands. Research thus far completed uniformly provides support for the behavioral properties ascribed to the need for social approval variable. Of direct relevance to the present study is the finding of Crowne and Strickland (1961) that individuals with a strong need for social approval emit significantly more plural nouns in a verbal conditioning situation than individuals less strongly motivated for social approval. The present study was designed to extend the findings of Crowne and Strickland to meaningful verbal communication in an interview situation. In accord with the view that individuals with a strong need for social approval are more sensitive and responsive to social reinforcers, it was hypothesized that individuals highly motivated for social approval would emit more responses in the reinforced category than individuals to whom social approval is of less importance. In addition, it was hypothesized that the high need for approval group would show a significant increase in the reinforced response class compared to their initial rate of emitting the reinforced responses. The latter hypothesis was prompted by Rogers' (1960) inability to produce a significant increase from operant to conditioned level in the amount of positive self-references in a quasitherapy setting similar to that of the present experiment.

## METHOD

### *Subjects*

A total of 76 undergraduate college students (34 males and 42 females) at the University of Kentucky volunteered to participate in the experiment. Sixty-four subjects were obtained in introductory psychology classes while 12 of the females were obtained from an introductory course in nursing. Subjects were requested to volunteer for a study of "the personal beliefs of college students."

### *Procedure*

The experimental procedure consisted of a single 15-minute individual interview with each subject. The interview was conducted immediately after the subject completed the Marlowe-Crowne Social Desirability scale which served as the measure of need for social approval. The M-C scale consists of 33

[1] This study was conducted while the author was at the University of Kentucky College of Medicine.

items with True and False response categories. An illustrative item is, "I never resent being asked to return a favor."

Each subject was told that the experimenter was making a study to learn what college students think and feel about themselves, and was instructed to describe his personality characteristics and traits. Following any silence of 20 seconds, the experimenter asked one of four questions depending on the content of the subject's verbalizations prior to that period of silence. In effect the subject was questioned about an area he had not previously discussed. The following four questions were used:

Could you describe yourself in terms of your relationships with people in general?
Could you describe yourself in terms of your relationships with members of your family?
Could you describe yourself in terms of your social life?
Could you describe yourself in terms of school work?

Every interview was tape recorded with the subject's knowledge, and later transcribed so that a content analysis and reliability check could be carried out from the complete record.

Subjects were assigned at random to either the experimental or control group without knowledge on the experimenter's part as to the subject's approval score. In the experimental group, every positive self-reference was reinforced by the experimenter's stating "Mm-hm" in a flat monotone. In the control group, the experimenter offered no reinforcement, and restricted his comments to asking one of the four questions when a 20-second period of silence occured. A positive self-reference was defined as any statement containing one or more references to subject personally as indicated by the use of such pronouns as "I," "My," "Me," "Mine," "We," etc., which served to describe subject in a favorable manner. A statement was defined as a sentence or complete "thought-unit" after Dollard and Mowrer (1947).

TABLE 1

MEAN PSR RATIOS AND STANDARD DEVIATIONS BY PERIOD FOR THE HIGH AND LOW NEED APPROVAL GROUPS UNDER EXPERIMENTAL AND CONTROL CONDITIONS

| Group | | Period 1 | 2 | 3 | 1–15 |
|---|---|---|---|---|---|
| Experimental | | | | | |
| High need approval ($N$ = 24) | $M$ | .181 | .233 | .166 | .194 |
| | $SD$ | .075 | .094 | .095 | .064 |
| Low need approval ($N$ = 14) | $M$ | .178 | .130 | .115 | .140 |
| | $SD$ | .086 | .079 | .058 | .062 |
| Control | | | | | |
| High need approval ($N$ = 18) | $M$ | .229 | .121 | .124 | .155 |
| | $SD$ | .117 | .082 | .083 | .063 |
| Low need approval ($N$ = 20) | $M$ | .203 | .114 | .075 | .129 |
| | $SD$ | .082 | .067 | .019 | .045 |

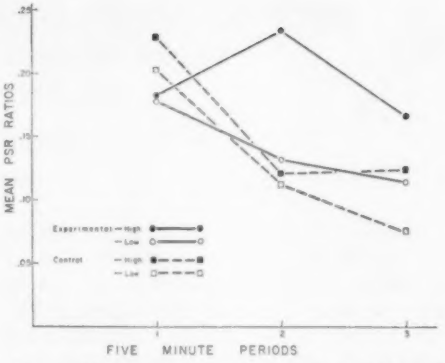


FIG. 1. Mean PSR ratios by period for high and low need approval groups under experimental and control conditions.

Immediately following the completion of the 15-minute interview, each subject was questioned to ascertain the extent to which he was aware of the conditioning procedure. The following questions were used in this part of the experiment:

What do you think is the purpose of the experiment?
Why did you say what you did?
Did you pay any attention to me while you were talking?
Did you notice anything I did while you were talking?
Why do you think I did that?
Did it have any effect on what you did?
What was the effect?

Reliability for the scoring of positive self-references was established by correlating the ratings of the experimenter with those independently obtained by another scorer who worked from the typed transcripts with experimenter's reinforcements omitted. A Pearson $r$ of .89 was obtained based on a sample of 15 randomly selected interviews.

## RESULTS

As an initial step in the analysis of the data, subjects in the experimental and control groups were dichotomized at the overall mean M-C score (15.36) to yield the high and low need for approval groups. Male and female subjects in each of the groups did not differ significantly with respect to mean M-C scores or rate of emitting positive self-references. The final analysis of the data was carried out, therefore, on the pooled sample of males and females.

The basic response unit used in the analysis is the ratio of positive self-reference (PSR)

statements to total statements made. The 15-minute interview was divided into three 5-minute periods. Mean PSR ratios for each 5-minute period and for the entire 15 minutes were computed for the high and low need for social approval subjects in the experimental and control groups. Table 1 contains the results of this analysis. The mean PSR ratios for the 5-minute periods are graphically illustrated in Figure 1.

In order to test the hypothesis that high need for social approval subjects produce more positive self-references under positive reinforcement than low need for approval subjects under positive reinforcement, $t$ tests of the differences in mean PSR ratios between the relevant groups were computed. For the first 5-minute period, no significant differences were obtained between the mean PSR ratios of any two groups. The maximum difference (that between the control highs and experimental lows) yielded a $t$ value[2] of 1.35, $p = > .05$. It would appear then that during the first period, the conditioning effect had not yet begun to differentiate between the groups. The mean PSR ratio of .233 obtained by the high approval subjects under positive reinforcement during the second period is significantly higher ($p = < .01$) than the corresponding mean ratios of each of the other three groups. For the last 5-minute period, the PSR ratio of .166 obtained by the high need for approval subjects under positive reinforcement is significantly greater than that of the experimental lows ($t = 1.82$, $p = < .05$), and the control lows ($t = 3.65$, $p = < .01$), although not significantly different ($p = > .05$) from the PSR ratio of the control highs. When mean PSR ratios for the entire 15 minutes are compared, the high need for approval subjects in the experimental condition are significantly ($p = < .05$) above each of the three comparison groups in rate of emitting positive self-references. Over-all, it may be concluded, as predicted, that individuals with a high need for social approval emit more positive self-references under positive reinforcement than subjects less strongly motivated for social approval.

[2] One-tailed tests of significance were computed throughout the analysis since directional hypotheses were formulated.

That the differences between the high need for approval subjects in the experimental condition and the three comparison groups are attributable to a conditioning effect, and not merely to a decline in the PSR ratios of the other groups is indicated by the significant increase ($t = 2.48$, $p = < .05$) in the PSR ratio of the high approval subjects from the first to second period. In contrast both control groups evidenced a significant decline ($p = < .01$) in PSR ratios from the first to the second period. The decline from the first to second period for the experimental lows is also significant ($t = 2.00$, $p = < .05$). The significant decline ($p = < .01$) in the PSR ratio of the high approval experimental subjects during the third period may be likened to an "exhaustion" phenomenon. This finding is similar to the "exhaustion" of plural noun responses obtained by Crowne and Strickland (1961) during the later half of their experimental session.

For the purposes of the present study, "awareness" was defined as an explicit statement by subject that experimenter said "Mm-hm" each time subject made a positive self-reference. None of the experimental subjects were aware of the "response-reinforcement" contingency, a finding similar to those of other studies (e.g., Krasner, 1958; Rogers, 1960). Of the 38 experimental subjects, 27 reported that they noticed the "Mm-hm." In general, these 27 subjects stated that the sound signified "interest" on experimenter's part or "encouragement" to continue talking. It may be concluded that changes in the rate of emitting positive self-references ocurred in the absence of any awareness on subject's part as to the nature of the experimental manipulations.

## DISCUSSION

The results of this experiment add considerable generality to the frequently reported finding that human verbal behavior is subject to control through the use of operant conditioning techniques. Previous interpretations of such findings with respect to their implications for psychotherapy and interviewing have been mitigated by the fact that the task typically posed the subjects in past studies have seldom required any personally meaning-

ful verbal communication. The present finding, that the frequency with which an individual makes positive self-references in an interview situation can be modified by minimal social reinforcements without awareness on subject's part, would appear to support Krasner's (1958) contention that process research in psychotherapy might profitably employ operant conditioning techniques. Conceivably, the change in patient's verbalizations in therapy from negative to positive self-references as therapy progresses may reflect to some extent, a subtle conditioning effect rather than improvement in the patient's psychological status.

The major findings of the present study indicate that individual differences in conditionability are in part attributable to differences in the strength of the need for social approval. As predicted, subjects in the high need for social approval group were more conditionable as evidenced by the significant increase in the rate of responding between the first and second periods. In contrast, the low need approval group failed to adopt responses reinforced by experimenter, as indicated by the significant decline in the performance of this group. Similarly, the two control groups evidenced a marked decline in rate of emitting positive self-references. These findings provide a clear-cut confirmation of data recently obtained by Crowne and Strickland (1961) in a study of the influence of need for social approval on the operant conditioning of plural-noun responses. Overall, the present results lend additional support to the construct validity of the need for approval variable. Previous research with this variable (Crowne & Marlowe, 1960; Marlowe & Crowne, 1961; Strickland & Crowne, in press) indicates that individuals characterized by strong tendencies to do and say the socially appropriate thing in order to obtain approval are likely to be conforming to external pressure, and to express attitudes congruent with perceived situational demands. In this study, the behavior of the high need for approval subjects may likewise be interpreted as reflecting the greater sensitivity and responsiveness of these subjects to social reinforcers as compared to individuals less strongly motivated to seek social approval. Instructions to the subjects were deliberately brief and nonspecific, requesting only that subject "describe his personality characteristics and traits." Those individuals who are most concerned with saying the socially appropriate thing were most likely to adopt the reinforced response. It should be noted that only a minor portion of the variance in responding with positive self-references is attributable to the influence of need for social approval alone. In the control group, which received no reinforcement, an insignificant Pearson $r$ of .23 ($N = 38$) was obtained between need for approval scores and the PSR ratios over 15 minutes. For the experimental subjects, a corresponding $r$ of .42 ($p = < .01$, $N = 38$) was obtained. Thus, the increase in productivity for the reinforced high need for approval subjects would seem to reflect their greater responsiveness to the reinforcements rather than any prior existing tendency to emit positive self-references.

The rather striking and significant decline among the experimental highs in the rate of emitting positive self-references during the third period might reflect a simple exhaustion phenomenon in that subjects ultimately run out of positive things to say about themselves. An alternative and probably more likely explanation is that the decline is due to a reluctance on the part of subjects to enumerate repeatedly their positive personal characteristics. Thus, one high need for approval subject who participated in the experimental condition spontaneously commented at the end of the interview, "It's awfully hard to sit here and toot your own horn." In line with the hypothesis that subjects are self-conscious about excessive self-praise is the comment by Matarazzo, Saslow, and Pareis (1960) that an individual's pre-experimental history of differental reinforcement to particular verbal response classes may exert an influence in verbal conditioning experiments. It seems quite likely that most individuals have developed the expectancy that excessive elaboration of one's good points is not socially acceptable behavior. The decline noted above, therefore, might well be due to the individual's prior reinforcement history exerting an effect against the specific reinforcements of the experimenter once a certain point has been

reached in the subjects production of positive self-references.

The present results regarding the conditionability of positive self-references are in conflict with those of Rogers (1960) who failed to obtain direct evidence of a conditioning effect over a series of six 10-minute quasi-therapy sessions. This discrepancy in results may be due to the fact that the intervals between sessions were of such length as to result in extinction of the reinforced responses from session to session. This point can only be considered suggestive, since Rogers does not specify in his report the number of days which intervened between each session.

## SUMMARY

The purpose of this study was to test the hypothesis that the conditionability of positive self-references in an interview setting is related to individual differences in the strength of the need for social approval.

Seventy-six subjects completed the Marlowe-Crowne Social Desirability scale which served as the measure of need for social approval and were then individually interviewed for 15 minutes under either control or experimental conditions. Experimental subjects were reinforced by the experimenter's stating "Mm-hm" each time subjects verbalized a positive self-reference. Control subjects received no reinforcement. Finally, each experimental subject was questioned to determine the extent of his awareness of the experimental manipulations. Each 15-minute interview was tape recorded and later typed. High interscorer reliability was obtained for the scoring of positive self-references.

The results obtained indicate that the rate of emitting positive self-references in an interview situation can be altered by operant conditioning, and that subjects with a strong need for social approval produce significantly more positive self-references under positive reinforcement than a comparable group of low need for approval subjects.

Results were interpreted in terms of their relevancy to process research in psychotherapy, and were further seen as contributing to the construct validity of the Marlowe-Crowne *SD* scale.

## REFERENCES

CROWNE, D. P., & MARLOWE, D. A new scale of social desirability independent of psychopathology. *J. consult. Psychol.*, 1960, **24**, 349–354.

CROWNE, D. P., & STRICKLAND, BONNIE R. The conditioning of verbal behavior as a function of the need for social approval. *J. abnorm. soc. Psychol.*, 1961, **63**, 395–401.

DOLLARD, J., & MOWRER, O. H. A method of measuring tension in written documents. *J. abnorm. soc. Psychol.*, 1947, **42**, 3–32.

GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409–416.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

MARLOWE, D., & CROWNE, D. P. Social desirability and response to perceived situational demands. *J. consult. Psychol.*, 1961, **25**, 109–115.

MATARAZZO, J. D., SASLOW, G., & PAREIS, E. N. Verbal conditioning of two response classes: Some methodological considerations. *J. abnorm. soc. Psychol.*, 1960, **61**, 190–206.

ROGERS, J. M. Operant conditioning in a quasi-therapy setting. *J. abnorm. soc. Psychol.*, 1960, **60**, 247–252.

SALZINGER, K. Experimental manipulation of verbal behavior: A review. *J. gen. Psychol.*, 1959, **61**, 65–94.

STRICKLAND, BONNIE R., & CROWNE, D. P. Conformity under conditions of simulated group pressure as a function of the need for social approval. *J. soc. Psychol.*, in press.

TAFFEL, C. Anxiety and the conditioning of verbal behavior *J. abnorm. soc. Psychol.*, 1955, **51**, 496–501.

(Received January 6, 1961)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 84–89

# AGGRESSION AND HOSTILITY IN PSYCHIATRIC PATIENTS[1]

ARNOLD H. BUSS
*University of Pittsburgh*
HERBERT FISCHER
*Purdue University*
AND ALVIN J. SIMMONS
*University of Massachusetts*

This study was designed to yield three kinds of information. First, it sought to establish the relationship between different sources of information about patients' aggressiveness. If the relationships are strong, aggressiveness is sufficiently pervasive to be assessed by any one of several means. If the relationships are weak, aggressiveness is too complex and varied to be assessed by a single source of information. Second, it related enduring aggressive trends to the way hostile materials are dealt with in the laboratory. Previous research has had a similar aim, but this is the first study to employ varied personality measures of aggression and varied laboratory tasks. Furthermore, aggressiveness is divided into different modes (assault, indirect, etc.), and each mode is related to performance with hostile materials. Third, the study sought further information on how aggressive behaviors tend to cluster. The Buss-Durkee inventory (Buss & Durkee, 1957) yielded factors of Aggression and Hostility when the scores of college students were factor analyzed. In the present study the scores of patients on this inventory are factor analyzed, as well as ratings from other sources of information. These factor analyses should indicate the stability and generality of the two factors derived previously.

## METHOD

### *Subjects*

The subjects were admissions to an acute treatment center for psychiatric patients. The age range was 16 to 60 years, and the diagnoses included neurosis, psychosis, character disorders, and mixed diagnoses. There were originally 100 subjects, but 4 were subsequently discovered to have organic brain damage, and they were dropped from the study. Of the remaining 96 subjects there were 63 women and 33 men, reflecting the approximately two-to-one ratio of women to men admitted to the hospital. Whenever there were clear sex differences in the data, the men's and women's results were treated separately.

### *Measures*

*Ratings.* Each patient's psychiatrist rated him on a seven-point scale for the variables tapped by the inventory: Assault, Indirect, Irritability, Negativism, Resentment, Suspicion, and Verbal. The psychiatrist made his ratings after studying the patient's social and psychiatric history and after at least one lengthy interview; test results were not available.

Each patient's psychologist made similar ratings, using only his observations of the patient's test behavior and the test protocols.

The closest available relative was interviewed. Because relatives are biased and have little experience with ratings, they were not allowed to make the ratings. Rather, one investigator interviewed and later read the protocol to the other two investigators, and then all three made independent ratings. Their intercorrelations ranged from .83 to .96. Because of the close agreement the three ratings for each category were pooled.

*Tests.* The Buss-Durkee inventory was administered to each patient individually. The Rorschach was also administered individually, and the Elizur (1949) and DeVos (1952) systems were used to obtain a hostile content score. The two systems were found to correlate .74, which is very high in light of the reliability of each system. Therefore only the Elizur system was used.

The Iowa Picture Interpretation Test (IPIT) (Hurley, 1955) was also administered individually. The form used here consisted of 24 TAT cards, with four alternative stories for each picture. The subject rated each story for its applicability to the picture, and the aggression score was the sum of the ratings for the aggressive stories. Because low scores equal high aggressiveness and vice versa, the correlations between the IPIT and other measures were reversed in sign.

*Laboratory tasks.* The scrambled sentences technique, first reported by Watson, Pritzker, and Madi-

[1] Analysis of the data was supported by a research grant, M-2655 (A), from the National Institute of Mental Health, United States Public Health Service.

TABLE 1

CORRELATIONS BETWEEN PSYCHOLOGISTS', PSYCHIATRISTS', AND RELATIVES' RATINGS

| Mode of aggression | Psychologist vs. psychiatrist (63) | Psychologist vs. relatives (64) | Psychiatrist vs. relatives (43) |
|---|---|---|---|
| Assault | .04 | .39** | .60** |
| Indirect | .30* | .08 | .20 |
| Irritability | .68** | .02 | −.01 |
| Negativism | .28* | .18 | −.13 |
| Resentment | .32* | −.09 | .28 |
| Suspicion | .46** | .17 | .30* |
| Verbal | .07 | −.12 | .11 |
| Sum | .31* | .18 | .18 |

* Significant at .05 level.
** Significant at .01 level.

son (1955), involves a neutral or hostile solution in unscrambling four words. Thus "take arm his break" can be unscrambled to form "take his arm" or "break his arm" when one of the four words is omitted. Some of the Watson et al., sentences were retained, but most were discarded in favor of newly constructed sentences. Thirty sets of words were used, and the score was the number of hostile solutions.

The operant conditioning task was a modification of that originated by Buss and Durkee (1958). On each trial the subject was handed a $3 \times 5$ card containing a hostile and a neutral word and was instructed to make up a sentence using one of the words. There were 60 trials in all. On the first 10 trials the subject made up the sentences without reinforcement; subsequently the experimenter said "Good" after each hostile response.

## RESULTS

The results will be reviewed under three headings. First, the relationships between the inventory and all other measures and tasks will be discussed. Next, the relationships of the ratings, tests, and laboratory tasks will be reported. Finally, factor analyses of the inventory and the ratings will be reviewed.

### *The Inventory*

The inventory scores were correlated with each of the other types of measures: ratings, tests, and laboratory tasks.

*Ratings.* Before inquiring about the relationship between these three sets of ratings and the inventory, it is important to establish the relationships among the ratings themselves. The relevant data are presented in Table 1. Psychologists' ratings were available on 95 of the 96 subjects, psychiatrists' ratings on 64 subjects, and relatives' ratings on 67 subjects. Thus the various sets of correlations differ in $N$, and the numbers in parentheses indicate these $N$s.

Concerning the correlations, six of the eight psychologist-psychiatrist correlations are significant, while in the remaining two columns only 3 of 16 correlations are significant. There is more agreement between psychologists and psychiatrists than between relatives and either psychologists or psychiatrists.

There appear to be two reasons for the low correlations among ratings. First, there were many raters and sources of information, and it may be assumed that each had his own implicit norms. Differences in such norms would lead to differences in ratings that would appear as unreliability; the result would be lower correlations.

Second, there are probably differences in aggressiveness from one context to the next, e.g., a patient behaves differently at home from the way he does in the hospital. Such discrepancies in behavior would lead to disagreement among raters who evaluate the patient's aggressiveness in different contexts. This has consequences for the relationship between self-report measures of aggressiveness and the observations of others. If the patient differs in aggressiveness in different situations, this places a low ceiling on the relationship between the inventory and the ratings.

The correlations between the inventory and the ratings are presented in Table 2. Only one of the psychological ratings was

significantly related to an inventory scale: Resentment. Since none of the other correlations in the Psychologist column is as high as .10, the Resentment correlation may be attributed to chance.

The psychiatric ratings, on the other hand, are related to inventory scores; four of the seven inventory scales are significantly related to psychiatric ratings. Evidently the vantage point of psychiatrists was better than that of psychologists, at least in relation to the patient's self-reports. The psychiatrists observed patients during interviews, listened to their stories, and had access to case history materials; they may have placed considerable weight on the patients' own stories.

While neither the psychologist-inventory nor the psychiatrist-inventory correlations showed sex differences, there were sex differences in the relative-inventory correlations. Therefore the men's and women's correlations are presented separately in the last two columns of Table 2. None of the men's correlations is significant; the women's correlations are for the most part significant but not high.

*Tests.* The correlations between the tests and the inventory are presented in Table 3. Again there are sex differences; for men Elizur hostile content scores are unrelated to inventory scores, but for women four of the inventory scores are related to the Elizur. For the IPIT the sex difference is in the opposite direction: significant correlations for men but not for women. The correlations between the IPIT and the Elizur hostile content score are .18 for men and −.40 for women.

TABLE 2

CORRELATIONS BETWEEN THE INVENTORY AND THE RATINGS

| Mode of aggression | Inventory versus Psychologist (95) | Psychiatrist (64) | Relative Men (22) | Relative Women (45) |
|---|---|---|---|---|
| Assault | .04 | .22 | .03 | .14 |
| Indirect | −.09 | .14 | .13 | .31* |
| Irritability | .06 | .36** | .03 | .29* |
| Negativism | .05 | .49** | −.30 | .24 |
| Resentment | .35** | .28* | −.30 | .36* |
| Suspicion | .04 | .20 | .16 | .30* |
| Verbal | .03 | .48** | .07 | .30* |
| Sum | −.03 | .19 | −.17 | .33* |

* Significant at .05 level.
** Significant at .01 level.

TABLE 3

CORRELATIONS BETWEEN THE INVENTORY AND ELIZUR HOSTILE CONTENT SCORES AND THE IPIT

| Mode of aggression | Inventory versus Elizur Men (33) | Elizur Women (63) | IPIT Men (27) | IPIT Women (50) |
|---|---|---|---|---|
| Assault | .19 | .26* | .40* | .18 |
| Indirect | .21 | .23 | .28 | .00 |
| Irritability | .01 | .17 | .11 | −.01 |
| Negativism | .15 | .20 | .36 | .21 |
| Resentment | .02 | .20 | .45* | .16 |
| Suspicion | −.01 | .30* | .49* | −.24 |
| Verbal | .12 | .37** | .41* | −.14 |
| Sum | .10 | .41** | .42* | .10 |

* Significant at .05 level.
** Significant at .01 level.

*Laboratory tasks.* The laboratory tasks were scrambled sentences and operant conditioning of hostile verbs. There were two scores for the latter task, operant level and a learning score. The learning score yielded no significant correlations with the inventory or any other measure, and it is omitted from further comparisons. The results for the scrambled sentences and operant level are presented in Table 4. For men one scrambled sentences-inventory correlation is significant, while for women all but one are significant. The correlations between the inventory and operant level show just the opposite results: fairly high correlations for men and very low correlations for women. The men's correlations between the inventory and operant level were among the highest of the entire study. The sex differences carried through to the scores on scrambled sentences and operant level, in both instances men scoring higher than women. The difference is significant at the .05 level for operant level ($t = 2.02$, $df = 64$) but just misses the .05 level for scrambled sentences ($t = 1.89$, $df = 88$).

### *Ratings, Tests, and Laboratory Tasks*

The three sets of ratings were correlated with the Elizur hostile content score, IPIT, operant level, and scrambled sentences. Since there were 21 ratings (three sets of seven

each) and each rating was correlated with four scores (two tasks, two tests), there were 84 correlations in all. Approximately four of these correlations would be significantly different from zero at the .05 level by chance alone, and nine correlations were actually significant at the .05 level. Since there is no way of knowing which correlations are reliably different from zero, the only recourse is to search for consistency. If a particular scale had significant correlations with several variables, its relationships are more likely to be stable. Only two scales manifested consistency: Irritability and Negativism, as rated by relatives, significantly correlated with the IPIT, scrambled sentences, and operant level. Thus patients' Irritability and Negativism prior to hospitalization is related to measures of aggressiveness.

Next the tests were correlated with the laboratory tasks, with the results kept separate. None of the correlations between the Elizur hostile content scores and the two laboratory tasks is significant. The IPIT-scrambled sentences correlations are .60 for men and .35 for women. The IPIT-operant level correlations are .66 for men and .35 for women.

The scrambled sentences-operant level correlations are .65 for men and .29 for women. The communality among the three measures is strong for men, and at least a weak relationship is present for women. The correlations between the IPIT, scrambled sentences, and operant level may be explained by assuming that each represents a base (or operant) level of aggressive responding. On the IPIT the subject chose from four stories to each picture, with one of the choices being hostile. On scrambled sentences the subject chose between a hostile and a neutral unscrambling of the sentence. On the learning task (before reinforcement) the subject chose a hostile or a neutral verb in constructing a sentence. Thus on all three tasks the subject chose between neutral and hostile alternatives, his score representing a base level of aggressive or hostile responding.

TABLE 4

CORRELATIONS BETWEEN THE INVENTORY AND SCRAMBLED SENTENCES AND OPERANT LEVEL

| Mode of aggression | Inventory versus: Scrambled sentences, Men (28) | Scrambled sentences, Women (52) | Operant level, Men (25) | Operant level, Women (41) |
|---|---|---|---|---|
| Assault | .40* | .40** | .49* | .14 |
| Indirect | .16 | .31* | .43* | .01 |
| Irritability | .04 | .16 | .38* | −.18 |
| Negativism | .06 | .36** | .65** | .01 |
| Resentment | .25 | .35* | .63** | .10 |
| Suspicion | .24 | .39** | .69** | .23 |
| Verbal | .19 | .34* | .56** | .02 |
| Sum | .22 | .43** | .63** | .09 |

* Significant at .05 level.
** Significant at .01 level.

TABLE 5

FACTOR LOADINGS FOR PATIENTS' INVENTORY SCORES

| Mode of aggression | Men (53) I | Men (53) II | Women (114) I | Women (114) II |
|---|---|---|---|---|
| Assault | .28 | .61 | .20 | .61 |
| Indirect | .35 | .63 | .07 | .30 |
| Irritability | .63 | .53 | .79 | .23 |
| Negativism | .34 | .68 | .29 | .55 |
| Resentment | .63 | .51 | .59 | .34 |
| Suspicion | .52 | .63 | .35 | .46 |
| Verbal | .33 | .69 | .43 | .59 |

*Factor Analyses*

Inventory scores were available not only for the 96 patients of the project but for 71 additional psychiatric patients; with the large number of subjects it was possible to keep the sexes separate. For the psychiatrists', psychologists', and relatives' ratings separate analyses were not necessary because the sex differences in correlations were small and unsystematic.

The 650 computer was used, and factors were rotated to an orthogonal solution by the varimax method. In order to make all the data comparable, the original inventory data on 85 college men and 88 college women (see Buss & Durkee, 1957) were again factor analyzed and rotated by the varimax method. The factor patterns are identical to those of the previous study: Factor I is defined by Assault, Indirect, Irritability, and Verbal; and Factor II is defined by Resentment and Suspicion. This pattern holds for both sexes,

TABLE 6

Factor Loadings of Psychiatrists', Psychologists', and Relatives' Ratings

| Mode of aggression | Psychiatrists (69) | | Psychologists (95) | | Relatives (67) | |
|---|---|---|---|---|---|---|
| | I | II | I | II | I | II |
| Assault | .55 | −.07 | .48 | .23 | .54 | .02 |
| Indirect | .71 | .26 | .54 | .33 | .84 | .00 |
| Irritability | .81 | .16 | .81 | .14 | .77 | .32 |
| Negativism | .68 | .17 | .10 | .61 | .60 | .38 |
| Resentment | .39 | .49 | .25 | .68 | .77 | .07 |
| Suspicion | .32 | .59 | .29 | .56 | .29 | .40 |
| Verbal | .82 | .17 | .68 | .22 | .82 | .04 |

but for women Negativism loads up on Factor I.

The inventory data for patients are presented in Table 5. The criterion of meaningfulness is the same as in the previous study: loadings of .40 and above. It is difficult to discern a pattern in the loadings above .40. For men Factor I is defined by Irritability, Resentment, and Suspicion; for women it is defined by Irritability, Resentment, and Verbal. For men all the scales have loadings greater than .40 on Factor II, whereas for women only four scales have loadings greater than .40. Thus for the men Factor II is a general factor, including all aspects of aggression and hostility, whereas for the women there is no general factor. However, the women's pattern for Factor II does not seem to make much psychological sense, nor does it conform to the general factor of the men. In brief, the factor analyses of the patients' inventories yield a confusing picture that suggests no clear interpretation.

The factor analyses of the ratings by psychiatrists, psychologists, and relatives are presented in Table 6. For psychiatrists, Factor I is defined by Assault, Indirect, Irritability, Negativism, and Verbal; Factor II is defined by Resentment and Suspicion. This pattern is identical to the pattern of college women taking the inventory. For psychologists, Factor I is defined by Assault, Indirect, Irritability, and Verbal; Factor II is defined by Negativism, Resentment, and Suspicion. This pattern is identical to that of the psychiatrists' ratings, with the exception of Negativism, which now loads up on Factor II. Thus except for Negativism, the factor pattern for psychologists' ratings resembles that for both psychiatrist's ratings of patients and college students' inventories.

Concerning relatives, Factor I is defined by all the variables except Suspicion, which defines Factor II. This factor pattern differs from the others in that Resentment now loads up on Factor I, whereas in the patterns for psychiatrists, psychologists, and college students Resentment combined with Suspicion to define Factor II.

These various factor patterns reveal both inconsistencies and consistencies. First, Negativism appears in Factor I, Factor II, or neither, depending upon the source of information; Negativism is an unstable scale. Second, the factor analyses of the patients' inventories yield no discernible pattern. Third, with the exception of the patients' inventories, all sources of information yield a factor pattern that is more or less consistent:

| AGGRESSIVENESS | HOSTILITY |
|---|---|
| Assault | Resentment |
| Indirect | Suspicion |
| Irritability | |
| Verbal | |

The two factors, Aggressiveness and Hostility, account for the loadings over .40, with the exceptions noted. They show stability across sources of information: self-reports of normals and ratings of patients by psychiatrists, psychologists, and relatives all yield more or less the same two factors. The factor analyses of patients' inventories did not yield the same factorial structure, which may be attributed to distortions by the patients in describing their own behavior. Thus the evi-

dence for the factors of Aggressiveness and Hostility is clearly positive.

## SUMMARY AND CONCLUSIONS

This study yielded several salient findings. The ratings by psychologists, psychiatrists, and relatives tended to be unrelated or only weakly related perhaps because Aggressiveness and Hostility are not pervasive personality characteristics that can be assessed with consistency across situations. A network of empirical relationships was established between the Buss-Durkee inventory and psychiatric ratings, relatives' ratings, projective measures of aggression, and laboratory tasks involving hostile materials. Both tests (IPIT and Elizur) were found to correlate with the inventory. The IPIT yielded higher correlations, and in addition it was significantly related to both laboratory tasks, whereas the Elizur was not.

Conditioning of hostile material was unrelated to any of the other measures used in the study, but the operant level of emitting hostile verbs was found to correlate with a number of measures and was one of the best measures of the entire study. The scrambled sentences were also related to several other measures. Both these laboratory tasks tap some base (operant) level of emitting aggressive responses, and they may be better measures of aggressiveness than ratings made on the basis of interviews or a battery of tests.

Finally, the factor analyses yielded fairly consistent factor patterns. The patients' inventory scores yielded no meaningful patterns, but the psychiatrists', psychologists', and relatives' ratings all yielded an Aggressiveness factor and a Hostility factor.

## REFERENCES

BUSS, A. H., & DURKEE, ANN. An inventory for assessing different kinds of hostility. *J. consult. Psychol.*, 1957, **21**, 343–349.

BUSS, A. H., & DURKEE, ANN. Conditioning of hostile verbalizations in a situation resembling a clinical interview. *J. consult. Psychol.*, 1958, **22**, 415–418.

DEVOS, G. A quantitative approach to affective symbolism in Rorschach responses. *J. proj. Tech.*, 1952, **16**, 133–150.

ELIZUR, A. Content analysis of the Rorschach with regard to anxiety and hostility. *J. proj. Tech.*, 1949, **13**, 247–284.

HURLEY, J. R. The Iowa Picture Interpretation Test: A multiple choice version of the TAT. *J. consult. Psychol.*, 1955, **19**, 372–376.

WATSON, R. E., PRITZKER, L., & MADISON, P. Hostility in neurotics and normals. *J. abnorm. soc. Psychol.*, 1955, **50**, 36–40.

(Received January 11, 1961)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 90-98

# HIGHER INTELLECTUAL FUNCTIONS IN A PATIENT WITH HEMISPHERECTOMY FOR TUMORS[1]

JAN H. BRUELL AND GEORGE W. ALBEE[2]

*Western Reserve University*

The clinical literature contains a number of reports on the mental status of patients following cerebral hemispherectomy. In many cases clinicians have been struck by the remarkable preservation of higher mental functions following so radical an operation. The literature on the subject was summarized by Mensh, Schwartz, Matarazzo, and Matarazzo (1952), and by Gardner, Karnosh, McClure, and Gardner (1955). We had an opportunity to interview and to test one such patient and would like to report our findings and observations.

## CASE HISTORY

The patient, a 39-year-old white man, was admitted to Western Reserve University Hospitals on October 14, 1954. He had been in good health until August 1954, when he developed headaches which became progressively more severe.

On admission positive findings were confined to severe papilledema, and some loss of fine movements of the left hand. Skull films, electroencephalogram, and ventriculogram gave evidence of a right frontotemporal mass. On October 18, 1954, a craniotomy was performed and a tumor was removed from the right frontotemporal area. Histopathologic diagnosis of the tumor was astrocytoma, grade four (glioblastoma multiforme). The patient had an uneventful recovery and was discharged on October 30, 1954, without demonstrable neurologic effects, except for a left superior quadrant homonymous visual field defect.

The patient was readmitted on November 25, 1954, for a right hemispherectomy and, on this admission, the only change in his examination since discharge was a recurrence of his papilledema. On November 26, 1954, a right hemispherectomy was done. This included the globus pallidus, putamen, and caudate nucleus, and a considerable portion of the thalamus. The major arterial supply was ligated just distal to its connection to the circle of Willis.

Postoperatively the patient had a left hemiplegia, disturbance of body consciousness on the left with loss of localization of stimuli applied to the left side. During the first few days after the hemispherectomy the patient refused to acknowledge his paralyzed left side. When he was made to touch his left hand or leg he insisted the hand and leg belonged to someone else. This symptom, encountered frequently in patients after damage in the right cerebral hemisphere (Weinstein & Hahn, 1955), subsided after about 3 weeks. During the first 3 postoperative weeks the patient was deeply depressed, drowsy, and almost lethargic. Slowly, however, he started to take interest in his environment. He commenced physical therapy, and after 60 days of convalescence, he became ambulatory with a left lower leg brace. He was able to localize stimuli to the general areas of the left side. Body image was described normally and he was able to differentiate touch, vibration, and pain stimuli with 60% accuracy on the left, al-

[1] Research supported in part by Grant B 1041 from the National Institute of Neurological Diseases and Blindness, United States Public Health Service.

[2] The authors would like to thank Marguerite Hertz of Western Reserve University, Rochelle Wexler, formerly of King's County Hospital, and Gerald R. Pascal of the University of Tennessee for the permission to reproduce here their blind analyses of projective tests given to the patient. Because of space limitations we could present only excerpts from these analyses and thus we may have unwittingly distorted some of the interpretations of the projective material. The complete analyses, however, are available on request.

though he stated that all were variations of pain. On the seventy-first postoperative day, the patient was discharged.

The patient continued to improve as far as ambulation and self-sufficiency were concerned until 3 weeks before his last admission (April 10, 1955). At that time, he began to have pain in his right eye, with some loss of sensation over the first division of the right fifth nerve. Six days before admission, he began to have nausea and vomiting. On this admission, ventricular tap revealed increased intracranial pressure and increased protein (800 mg. percent). Right carotid arteriography revealed displacement of the left anterior cerebral and anterior communicating arteries, consistent with tumor regrowth. A ventriculogram revealed a space occupying mass in the right cranial vault. He was discharged without therapy and progressively deteriorated until his death on April 25, 1955. Unfortunately permission for a postmortem examination of the brain of the patient was not granted by his family.

## Observation of Behavior

Seventy-one days after the hemispherectomy the patient was discharged from the hospital. At home the patient apparently tried to overcome his depression by joking about his condition and making plans for his future. He felt that he could return to productive work in his former occupation as a machine designer. His wife reported that he did not show concern about their precarious financial situation. He did not ask how she managed to pay for household and hospital bills. On the contrary, he was puzzled as to why she had to spend endless hours in the kitchen making chocolates for a candy store in town, and he was annoyed that she did not spend all her time with him.

As he recuperated from the operation, he became increasingly preoccupied with sexual matters. On his return home he made frequent sexual advances to his wife but found himself impotent. He exposed his genitals freely, usually when he was alone in his chair, but he was quite unconcerned when someone entered the room. He also made sexual advances to a female neighbor who was helping his wife care for him. He felt very badly at "not being a man any more." He told his wife that since he could no longer have sexual relations with her, he might as well die.

During the last weeks of his life the patient showed on several occasions striking distortions of time sense. After his wife had just left the room she would be called back by the patient. The patient would cry and complain bitterly that he had been left alone and unattended for hours.

## Psychological Testing

We saw the patient for psychological testing during March of 1955 in the fourth month following hemispherectomy. Testing was done in his home, and because the patient tired easily, testing was spread over four afternoon sessions. Five subtests of the Wechsler-Bellevue Intelligence Scale (Form 1), the Rorschach, the Draw-a-Person Test, and the Bender Gestalt test were administered. The patient was extremely cooperative and eager to help us and science.

*Wechsler-Bellevue Intelligence Scale.* Unfortunately we were not able to give to the patient the complete test. On the five subtests we did administer the patient scored as follows:

| | RAW SCORE | WEIGHTED SCORE |
|---|---|---|
| Verbal scale: | | |
| Information | 14 | 10 |
| Similarities | 17 | 14 |
| Vocabulary | 33 | 14 |
| Performance scale: | | |
| Picture Completion | 7 | 6 |
| Object Assembly | 8 | 2 |

If we prorate for the missing subtests we arrive at a verbal IQ of 121, and a performance IQ of 76.

*Rorschach test.* The Rorschach test was completed in one session and was recorded on tape. The patient's performance on this test gives a better picture of his higher mental functions than any other observations we could contribute. We, therefore, reproduce the protocol in its full length.

Subject: R. G. (Male)
Age: 39
Date: March 22, 1955

I 3″

1. Well, that might be some coral formation. I have never seen any actually except in pictures. (The subject hands card back; the experimenter asks the subject to look some more to see if he see anything else there.)

(Q: You said it might be some coral formation. Something you might have seen in pictures. Can you tell me how much of the blot was the coral formation?)
S: This thing here—was not the coral formation. (When we get to it you tell me.)

2. Well, it could almost be a skull, but the eye is missing. (O.K.?) O.K.

60″

2. (You said it might be a skull.) Because of these two empty spaces, looks like the hollows of the eyes and the teeth that might have been across here. (But the whole thing was the skull?) Yeah. (If you see anything else when we go along you might tell me about it.)

II 9″

1. Looks something like a pelvis bone—but it isn't. (Long pause. Q: What else does it remind you of, anything?)

1. (You said something looked like a pelvis bone. What was that?)
These two halves here—the two sides of the pelvis. And this part here resembles the part where the spine is attached to it.

2. Can't think of anything.
(It does not have to be exact or perfect, just whatever it might remind you of.)
I am trying to correlate that red stuff in there but it doesn't ring any bell. The red throws it off altogether. (What would it be without the red?) Well, that black section looks something like a bat.

120″

Were those things created accidentally or by—drawn up?

2. (You said you couldn't correlate the red stuff. Does it look like anything to you now?) No.
(You said the black could look like a bat. Where is that?)
These two halves would be the wings. These two little points would be the hooks on the front of the wings and these are the feet back of the wings by which the thing hangs upside down.

III 2″

1. Well that red in the center could be a butterfly.

1. Because these look like butterfly wings but there is no true body in it. (Anything else make you think of a butterfly?)
Pause. (I was wondering if it were gray, would it still look like a butterfly?) Well, it's a pink coral color, not gray. (Do you think of a pink butterfly?) No, I don't think the color had much to do with it. Shape of it did, though.

2. And a sea horse in the other red.

2. (Q: Shape and color of sea horse.) Yes.

3. Well, this part is the same as in the other picture. (Q) That lower center section. This part here is the same as the other one there. (What does it look like or remind you of?) Well, a pelvis. Is that correct? (Um hum, anything else? Actually there are no real right or wrong answers). As long as it is a pelvis, this could be a femur. (Why is that? You say as long as that's a pelvis?) Well that's because that's the relationship. The bones. (Anything else?)

3. Looks like a pelvis—these two sections here, this to the center where the spine would come into it. Coccyx region.
(You said this could be a femur. Why?) Oh, that's the relationship to the pelvis. That would be one side or the other. There are two of these, you know. (Because of the relationship?) Yes. It's also long and thin like a femur, except a femur doesn't have this projection on it, normally, that is.

There is another red spot right here. It doesn't make any picture at all. Except the reverse of the opposite one. There is a sea horse on the other side. (Is this a sea horse too, then?) Yeah, this could be a sea horse. Maybe because the color is very much related. (How is that?) Well a sea horse is about the same color as that kind of a pink. (I remember seeing a sea horse in an aquarium, but it was pickled. Are they a pink color ordinarily?) Yes, coral color. Little deeper than a true pink. They cannot be kept alive in captivity. (Is that right?) If you had an aquarium they wouldn't live in it. You have seen the dead ones.

150″

IV 2″

1. Looks like a bearskin.
(Anything else?) That's all. (Hm) Can't see anything else there (the subject keeps card).

1. Well the shape of it resembles it and also the color. Dark gray-black sort of color. (And that makes it look like a skin?) Yeah.

2. This spot looks like a hippopotamus or a rhinoceros. (This part down here?) Yes, this head part here. D1
Seems to me I've seen these before somewhere.

2. (Then you said a hippo or rhino head?) These projections look like the horns on a rhinoceros or the general shape here is flat like a hippopotamus. There's a couple of eyes right here. Looking down on top of it.
The subject reacts to a humming noise and remarks: That sound you hear is a washing machine. It is not my heart.

50″

V 8″

1. Looks like a skinned rabbit. It's opened up and its ears sticking up and his feet stretched out behind him.

(You said this looked like a skinned rabbit with ears opened up. Can you tell me a little bit about how you see that?) Well, with the legs stretched out behind and the two ears up here and the other two legs here. (Q: opened?) Yes, split down the middle and then opened up and laid on its stomach side. (How can you tell?) The relation to the feet and head.

30″

VI 12″

It doesn't make any sense at all.
(Do you see anything here?)
No (the subject keeps card).
1. Except, maybe, it's a bear skin.
(Okay?) Okay.

1. Well there's a foot here, looks like the hind foot of a bear and this thing by a stretch of the imagination could be one of the front paws. These could be the ears here, or one of the ears. (You said on the other one something else beside shape. Is there anything else here?) Coloring resembles it.

60″

VII 4″

1. Um, could be a crab.
(Pause. Q: Remind you of anything else?) No. (Long pause. Q: You look like you are about to say something. Are you?)

1. (You said this could be a crab. Can you tell me how you saw that?) The body and the claws. (Here?) Correction please—this white part would be the body and these the claws. (Space in center is body and top are the claws. Q: Now, how are they attached? How much of this are claws?) This part right here. (Both sides, or one side?) Both sides. Actually, a crab has two more claws in the back. This one has lost his in a battle someplace but he'll grow new ones eventually.

2. I was going to say it has a faint resemblance to a flying squirrel but it actually doesn't.

2. (Then you said there was a faint resemblance to a flying squirrel.) Well, with these membranes stretched out, they look kind of thin here, like the skin would be stretched out. No such thing a flying squirrel. Actually it's a gliding device used to jump off one tree and glide to another. (And what made you think of that, this bottom part?) The thinning out of the colors here, looks like the membrane stretched out against the feet. (How much of the squirrel do you see?) Well if the feet were stretched out along here, which you don't see any feet of course. (It's really this shading that gives you the impression of skin being stretched.) Yes.

3. This part looks like a stabilizer on an airplane. (Bottom detail)

185″

3. The outline here. (Q) Cause it looks like a stabilizer on an airplane. The shape it's not a true shape of a stabilizer because it doesn't lend itself to an air foil section properly. (What is a stabilizer?) The wing section on the back of the plane (tail?) The appendage. It keeps the back end of the plane level with the front.

VIII 2″

1. Looks like a skeleton of some kind of fish. (D 3)

1. (You said you saw a skeleton of some kind of fish. Now where do you see that?) Well these little members across here could be some type of rib connections. This could be the main body of the skeleton. (The blue.) Yeah. This could be maybe the head or skull whatever you call it.

2. That red looks like a flower that is starting to open up—this part—not that part (Bottom center part?) Yeah.

2. Looks like the petals of a flower and the colors.

3. This part looks like that animal that shoots its tongue out at insects. (D 1)

3. This was an animal, can't think of its name you have seen it in movies probably, that shoots out its tongue against an insect to capture it. (Chameleon?) It's not a true chameleon. (Gila monster?) No, a gila monster is not as long as this. Different shape or color too. (Is this the right color?) No, it's strictly the shape of the thing.

4. This part here looks like the picture I've just seen. This section right in here. (You mean it looks like the last ink blot?) Yeah. (Just the middle part you mean here?) Yeah. (Doesn't it look like anything special?) No.

4. (Shows VII) This here is the same as this. (The blue here?) Very much like this bottom part in that card. (VII)

5. Well, it looks like a couple sails on a wind jammer. (Center blue)
(Okay?) Okay.

153″

5. (Couple of sails on a wind jammer, where are they?) (Blue) Right here. In technicolor. (Q) No, only in the movies. (So actually it's the shape?) Well, they make them a special color on those small star class sailing vessels. But these type of sails would not be suited for that. Square riggers is what they call them. (So it's mostly the shape here?) Yeah. The shape and the hollow—fading away of color here gives you the impression that it's full of wind.

IX 2″ Pretty colors 23″

1. Like the head of a moose. (Top projections)

1. (You said it looks like the head of a moose. What reminded you of that?) These things look like the antlers. Then there is a couple of eyes. (These little white spots here?) This here, grayish color in here could be the nose part of it.

2. This is the head of a person here—side view of him. With the head and the body. But this has no correlation at all (points to center bottom detail).

2. (Then you saw the head of a person. Sort of a side view) Right here is the head—here the nose, the mouth right in here. There is the body down here. It isn't a true body either. (What do you mean?) Well, it's too squat, not long enough. Unless it's a Japanese type of person, squatting down. (You said this has no correlation?) No.

67″

X 27″

1. This insect here—looks like an insect, what do you call them, black beetles? (Top detail)

1. Yeah. These things look like the pinchers of a beetle. These could be the legs dangling out back there (Q). He's attacking this thing whatever it is.

2. This again looks like a flower.

2. (Side yellow) This dark gray part looks like that section where the flower grown out of. Those are the petals. And the color helps.

3. This could be a sea worm of some type, who . . . . name I cannot say.

3. Then the sea worm. (Bottom green)

4. This could be some kind of a spider. (D 1)

4. All these little thin sections look like legs. The body right through there.

5. If this were a different color it could be a polar bear.

5. Don't forget the polar bear, if it was white! There's the head and there's a small tail on the bear. (And these are the feet?) Yeah, curled up underneath him. (You say they are curled up?) Yeah, up under his body.

6. This thing here looks like the weight of a centrifugal type governor. (Center detail)

6. These two balls are weights on the thing and it's rotating around this axis here and it opens as it whirls, it opens up it pulls up the center section through the shaft that governs the speed of the motor by opening and controlling the throttle on the carburetor, butterfly valve. (Spinning?) Well they are approaching, they are not quite out. Perhaps you remember the weights on the old phonographs. They use them on steam engines too, and auto engines. It isn't the most common type. The common type they use on truck engines is the air type. It keeps the engine from overspeeding, that's when an engine goes to hell, when it's overspeeding. It just knocks itself out. Vibration tears it down. That's why you should always keep an engine under load.

7. This thing could be a tree right here growing up. (Top)

8. If this body was thinner it would look like a praying mantis. It's too fat for that type of bug. How many legs does a praying mantis have? (Six) Then it's an insect. Well, let's look at it (Counts legs). This one doesn't have six legs. (Side brown)

8. These two prongs here on the front paws or claws, whatever you call them. These things up here are the legs holding it up. (Q) (Standing up?) Yeah. (Side gray)

230″

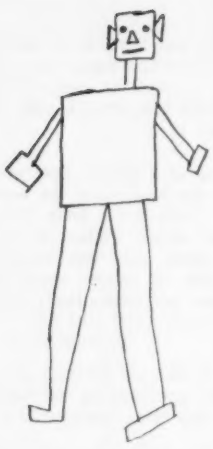


FIG. 1. Patient's drawing of a person.

*Drawn-a-Person test and Bender test.* The patient's drawing of a person and his reproductions of the forms of the Bender Gestalt Test are shown in Figures 1 and 2.

## BLIND ANALYSES OF PROJECTIVE TESTS

We were faced with the problem of obtaining an interpretation of the projective test material unbiased by a knowledge of the presence of such gross brain pathology. We, therefore, asked three authorities in their respective fields to undertake blind analyses of the test protocols. The only information given was that the subject was a 39-year-old male.

The patient's Rorschach protocol was surprising in many respects. It was longer and more complex than one would expect from an individual with such extensive brain loss. We are obliged to Marguerite Hertz for a detailed blind analysis of the Rorschach protocol. Only a few lines taken from Hertz's analysis shall be quoted here:

> The subject functions at a most mediocre level. While his thinking falls within the limits of accuracy there is evidence of considerable weakness in reality testing. Trends towards fluidity and arbitrariness of percepts and ideas appear too frequently in so short a record. The patient is concerned with bodily status, sexuality, potency. He seems filled with anxiety ridden fantasies of being torn, mutilated and/or deformed. Potentiality for feelings of unreality have been indicated. He shows strong potentiality for bizarre departures from realistic thinking and conventional logic. While subject may show a superficial semblance of being organized, he is schizophrenic-like in his handling of relationships and in his methods of coping with the complexities of reality. Paranoid features are likewise dominant in the record. Finally, there are many indications of organic impairment that should be investigated.

A blind analysis of the patient's drawing of a person was performed by Rochelle Wexler, and with her permission we present an excerpt from this analysis:

> Regarding his intelligence, his delineation of such sophisticated forms as the trapezoidal ears, as well as the whole "feel" of the drawing, suggests a man of good intelligence. He is so neat, so carefully organized, and his techniques can be carried out with such consistency that it would not surprise me to learn that he is working effectively. He achieves an almost too perfect mechanical integration. However, he achieves this integration in a highly arbitrary manner. He gets along in a mechanical, robot-like fashion. He has worked out his mode of adjust-

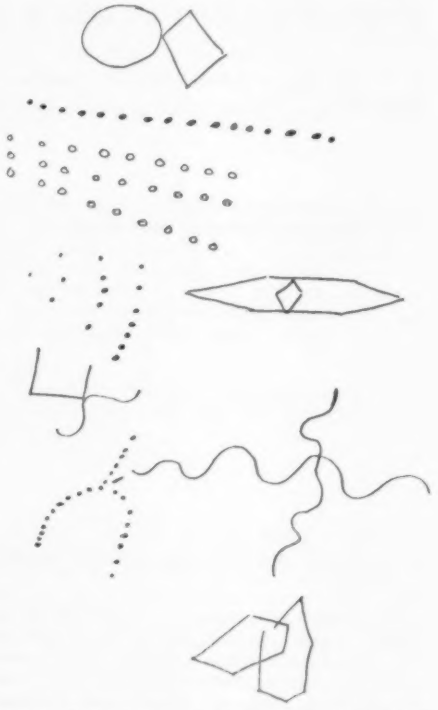

FIG. 2. Patient's reproductions of the forms of the Bender Gestalt test.

ment in such a rigid manner that any deviation would severely threaten his equilibrium; his defenses are so brittle, and his present adjustment is so tenuous that he could easily "go to pieces" under any stress.

Wexler notes many immature and childish features, major difficulties in the area of handling aggression, paranoid ideation, feelings of sexual impotence, tremendous feelings of personal inadequacy, depressive features. She feels that many features of the drawing "strongly suggest schizophrenia." However, "if he is schizophrenic, I would suspect that the pathology has an underlying organic basis."

The drawings of the Bender Gestalt Test were examined by Gerald R. Pascal. In his analysis, Pascal notes among other things "a stickiness and perseverative quality to Design 1 which may or may not be due to pervasive organic disorder" but then again he labels "organicity" as only a "remote possibility," and he feels that "we are dealing with an individual whose manifested behavior is mostly in contact but who is suffering from mood swings with underlying aggression, distrust, some suspiciousness of the environment, and whose obsessive-compulsive defenses seem to be no longer efficient. . . ."

## DISCUSSION

We met the patient in the fourth month following hemispherectomy. We were surprised to find a man who could discuss his condition in a seemingly detached and objective way. Our surprise was shared by most other people who met the patient. Had we had to base our judgment only on the personal contacts we had with the patient we would have stated that his higher intellectual functions survived hemispherectomy unimpaired. Examination of the test results, however, revealed a different picture. While some functions appeared remarkably preserved, others were markedly impaired.

Many authors have commented on the "concreteness" of brain injured patients. Such patients are said to live in a world of concrete objects rather than in a world of abstract ideas. Such concreteness was not obvious in this patient. His performance on the Similarities test was superior (IQ-equivalent 128). Most of his definitions on the Vocabulary test were abstract. Objects were defined in terms of their membership in a wider class rather than in descriptive terms.

Another indication of the "loss of the abstract attitude" is seen in the alleged inability of brain injured patients to plan their activities. If one considers the short term planning that can be observed in conversations or on psychological tests, loss of planning ability was not evident in this patient. Explanations the patient gave were orderly, one step in the explanation prepared the next step, and so on. An example of this is his explanation of the functioning of a centrifugal governor seen on Card X of the Rorschach. If, on the other hand, we consider long range planning, the patient showed signs of impairment. We refer to the observation reported above concerning the patient's disinterest in and lack of understanding of his financial situation, and also his lack of concern for the possible social consequences of his changed sexual behavior. He had become immodest in the presence of neighbors, and he did not seem to care whether they disapproved of it.

The perceptual functions of this patient had definitely suffered. Before his disease the patient had worked as a machine designer, and his last employer told us that the patient had done work which called for inventiveness and imagination. It appears inconceivable that at the time the patient designed machine parts, he would have performed as poorly on the Object Assembly test or the Picture Completion test as he did when we tested him. For example, in putting together the manikin of the Object Assembly test, the patient reversed the position of the two legs of the manikin without becoming aware of it, and on the Picture Completion test he identified the missing part only in 7 out of the 15 pictures. This failure cannot be attributed to the patient's hemianopsia as the pictures of this test are small enough to have been seen by the patient in their entirety. In the macular area patient's vision extended 2 to 3 degrees around the fixation point, and the patient had a corrected acuity of 20/20 in both eyes (Volk & Bruell, 1956). On the positive side, the patient's performance on the Bender Gestalt

Test must not be overlooked. The patient's drawings were not disorganized, and in the blind analysis of these drawings organicity was considered a "remote possibility" only. Thus, as in the area of abstract behavior, we have here this peculiar mixture of loss and preservation of functions.

The situation is similar when we consider the recall functions of the patient. On the one hand, we have the superior performance of the patient on the Vocabulary test (IQ equivalent 128), and the rich vocabulary displayed in the Rorschach protocol. By contrast, we find that on the Information test the patient scored just average (IQ equivalent 106). We were struck particularly by the lack of balance in the things he knew and did not know. Thus he would state that the airplane was invented by "Michelangelo and more recently by the Wright brothers," but he would estimate the population of the United States to be five million people. The patient's performance on the Rorschach also indicated some weakness in memory functions. He failed to recognize cards he had seen before, and then again he believed to have seen cards already which were shown to him for the first time.

Each of the three analysts of the projective test material considered the possibility of organic impairment, but each also noted features reminiscent of paranoia. The question arises whether the schizophrenia-like features of the patient's performance can be attributed to hemispherectomy. Unfortunately, this question cannot be answered with any degree of assuredness in the case of this particular patient. Close relatives of the patient including his mother and a brother had gone through schizophrenic episodes or had been hospitalized for schizophrenia, and there are also all indications that the patient's premorbid adjustment was, to say the least, schizoid. We would suggest that possibly the shock of the operation had activated in him latent schizophrenic adjustment tendencies but we would hesitate to relate schizophrenia-like features of his behavior to the removal of brain tissue. On the contrary, we would point out as remarkable that adjustment mechanisms mindful of schizophrenia which had existed preoperatively had survived hemispherectomy.

## SUMMARY

The case of a 39-year-old white male is reported whose right cerebral hemisphere was removed for tumor. Five subtests of the Wechsler-Bellevue Intelligence Scale, the Rorschach, the Draw-a-Person Test, and the Bender Gestalt Test were administered postoperatively. Test results are reported and discussed. The case confirms earlier findings of a remarkable, but by no means complete, preservation of higher mental functions following cerebral hemispherectomy for tumor. The patient showed a perfect retention of language, and he had no obvious difficulties in forming abstract concepts. Short term planning was preserved but there were signs of impairment of long range planning ability. There were also indications that the memory for remote and recent events had suffered. Perceptual functions were definitely altered. A marked loss of social inhibitions in the sexual adjustment of the patient was evident. Because of the schizoid premorbid adjustment of this patient, it was impossible to say whether some of the schizophrenic-like features in his postoperative behavior were a consequence of the operation or whether they persisted in spite of it.

## REFERENCES

GARDNER, W. J., KARNOSH, L. J., MCCLURE, C. C., JR., & GARDNER, A. K. Residual function following hemispherectomy for tumor and for infantile hemiplegia. *Brain*, 1955, **78**, 487–502.

MENSH, I. N., & SCHWARTZ, H. G., MATARAZZO, R. G., & MATARAZZO, J. D. Psychological functioning following cerebral hemispherectomy in man. *AMA Arch. Neurol. Psychiat.*, 1952, **67**, 787–796.

VOLK, D., & BRUELL, J. H. Eye movements in an adult with cerebral hemispherectomy. *Amer. J. Ophthal.*, 1956, **42**, 319–325.

WEINSTEIN, E. A., & HAHN, R. L. *Denial of illness: Symbolic and physiological aspects.* Springfield, Ill.: Charles C Thomas, 1955.

(Received January 18, 1961)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 99-100

# THE FIRST FACTOR OF THE MMPI:

## SOCIAL DESIRABILITY OR EGO STRENGTH?

ALLEN L. EDWARDS AND LOUISE B. HEATHERS

*University of Washington*

Kassebaum, Couch, and Slater (1959) have reported the results of a factor analysis of 32 MMPI scales. They interpret their first factor as "Ego-Strength vs. Ego-Weakness." No reference is made to Fordyce's (1956) interpretation of the first factor of the MMPI in terms of social desirability nor was Edwards' (1957) Social Desirability (*SD*) scale included in the analysis. The present study is concerned with the relationship between the first factor loadings of 31 of the 32 MMPI scales, as given by Kassebaum, Couch, and Slater, and the correlations of these scales with the *SD* scale.[1]

Correlations of 19 of the MMPI scales with the *SD* scale were available from other sources (Edwards, 1957, 1961). The remaining 12 correlations were obtained using the Merrill and Heathers (1956) sample of 155 males. Figure 1 shows the first factor loadings of the scales plotted against the correlations of the scales with the *SD* scale. The product-moment correlation between the factor loadings and the *SD* correlations is −.985. It is obvious, then, that the magnitude and sign of the first factor loadings of these MMPI scales could have been quite accurately predicted from the zero-order correlations of the scales with the *SD* scale.

To obtain a high score on the *SD* scale, a subject must give a relatively large number of socially desirable responses to the items comprising the scale. A socially desirable response is defined by Edwards (1957) as a True response to an item with a socially desirable scale value or a False response to an item with a socially undesirable scale value. Edwards, Heathers, and Fordyce (1960) and Edwards (1961) have shown that the correlation of an MMPI scale with the *SD* scale is a function of the percentage of items in the MMPI scale which are keyed for socially desirable responses. For the 31 scales of the present study, the product-moment correlation between these percentages and the first factor loadings is −.932.

To interpret the first factor as "Ego-Strength vs. Ego-Weakness" may be more satisfying from the viewpoint of personality dynamics, but, on the basis of the evidence we have reported, we believe it more reasonable to interpret the factor as "Social Desirability vs. Social Undesirability," that is, as representing the tendency to give socially desirable vs. the tendency to give socially undesirable responses to personality items of the True-False type. Regardless of which interpretation one chooses, the fact remains that the *SD* scale provides an objective and easily obtained measure of the factor.

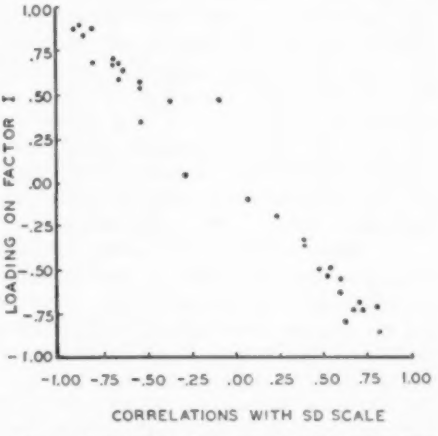


FIG. 1. Relationship between first factor loadings of 31 MMPI scales and correlations of the scales with the *SD* scale.

[1] A list of the scales is given by Kassebaum, Couch, and Slater. One of the 32 scales, the Social Presence scale, was not included in our analysis because of an apparent discrepancy in the scoring key used by Kassebaum, Couch, and Slater and the scoring key for this scale as given to us by Harrison Gough.

## REFERENCES

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

EDWARDS, A. L. Social desirability or acquiescence in the MMPI? A case study with the *SD* scale. *J. abnorm. soc. Psychol.,* 1961, **63**, 351–359.

EDWARDS, A. L., HEATHERS, LOUISE B., & FORDYCE, W. E. Correlations of new MMPI scales with Edwards SD scale. *J. clin. Psychol.,* 1960, **16**, 26–29.

FORDYCE, W. E. Social desirability in the MMPI. *J. consult. Psychol.,* 1956, **20**, 171–175.

KASSEBAUM, G. G., COUCH, A. S., & SLATER, P. E. The factorial dimensions of the MMPI. *J. consult. Psychol.,* 1959, **23**, 226–236.

MERRILL, R. M., & HEATHERS, LOUISE B. The relation of the MMPI to the Edwards Personal Preference Schedule on a college counseling center sample. *J. consult. Psychol.,* 1956, **20**, 310–314.

(Received November 28, 1960)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 101–102

# AGE AS A VARIABLE IN AVERSION TOWARD FOOD AND OCCUPATIONS

HENRY CLAY LINDGREN

*San Francisco State College*

We are all familiar with the person who "knows what he is against, but doesn't know what he is for." Such an attitude is usually considered characteristic of at least a minor degree of emotional disturbance. One aspect of this negativistic attitude was measured by Wallen (1948), among others, who found that servicemen discharged on psychiatric grounds reported a greater number of food aversions than did a control group.

It is the purpose of this paper to explore some of the relationships between age and negative choice and to show that the propensity toward making negative choices, albeit characteristic of certain neuropathological syndromes, is also characteristic of individuals in their late teens, as contrasted with adults in their twenties and older. The hypothesis to be tested is that younger people would reject more foods than older people, and that this age-related pattern of rejection would apply to occupations, as well as to foods.

Two questionnaires were prepared, the one consisting of a list of foods, and the other of occupations.[1] The food questionnaire included the 28 foods used in studies by Altus (1949) and Wallen (1948), supplemented by 11 others, selected because of their relatively rare appearance in the standard middle-class, American diet and added for the purpose of increasing the reliability of the questionnaire. The occupations consisted of 25 male and 26 female jobs, selected for their tediousness, low status, or unpleasant connotations or associations.

[1] Questionaires have been deposited with the American Documentation Institute. Order Document No. 6939 from ADI Auxiliary Publications Project, Photoduplications Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

The questionnaires were administered to 392 college students, who were asked to indicate which foods they disliked so much they would refuse to eat them, and which jobs (appropriate to their sex) they would not do under any circumstances. Scores were computed by counting the number of aversions. The reliability of the food questionnaire proved to be .84; reliabilties for the male and female portions of the occupations questionnaire were .84 and .86, respectively.

The mean number of food aversions for men 21 and under was 7.16, with a standard deviation of 4.12, while for men over 21, the mean was 4.41, with a standard deviation of 2.95. Women 21 and under had a mean of 6.72, standard deviation of 3.60; women over 21 had a mean of 4.45, standard deviation of 3.28. The differences between the mean food aversions reported by each age group were significant at the .001 level.

The mean number of occupation aversions for men 21 and under was 5.70, with a standard deviation of 4.78; for men over 21, the mean was 4.08, with a standard deviation of 4.07. The difference between means was significant at the .05 level. Means and standard deviations for women were 6.61 and 4.31 for women 21 and under, and 4.88 and 4.39 for women over 21, the difference between means being significant at the .01 level. Although women tended to reject more occupations than did men, this difference must be considered in light of the fact that women had a different list of occupations to check.

Presumably, if both food and occupation aversions were a function of a more generalized aversive attitude, scores made on the two questionnaires would be positively correlated. Product-moment correlation between the two questionnaires for the younger men was .48, significant at the .001 level; for the older men it was .22, significant at the .05 level; for the younger women it was .23, significant

at the .05 level; and for the older women it was .28, significant at the .01 level.

SUMMARY

College students 21 and under tended to indicate dislike for more foods and more occupations than did students over 21. The tendency to reject foods was positively correlated with the tendency to reject occupations.

REFERENCES

ALTUS, W. D. Adjustment and food aversions among Army illiterates. *J. consult. Psychol.*, 1949, **13**, 429–432.

WALLEN, R. Food aversions in behavior disorders. *J. consult. Psychol.* 1948, **12**, 310–312.

(Received January 10, 1961)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 103–104

# *BRIEF REPORTS*

## PSYCHOLOGICAL REACTIONS TO INFANTRY BASIC TRAINING [1]

PAUL EKMAN
*University of California School of Medicine*
WALLACE V. FRIESEN
*University of Kansas*
AND DANIEL R. LUTZKER
*Fort Dix, New Jersey*

One of the working assumptions of military psychiatry is that basic training has a beneficial psychological impact on the typical recruit. The plausibility of this hypothesis was explored by administering the MMPI to recruits selected from the first, fourth, and eighth weeks of basic training.

Ninety-three active army personnel having 12 years of education and between 18 and 22 years of age served as subjects. All of the subjects were drawn from three basic training companies at one military post. Three separate groups of recruits were tested, with men from each of the companies selected at each of the testing periods: 28 men were tested from companies in the first week of training, 34 different men from the fourth week, and 31 additional men from the eighth week.

The MMPI profiles without *K* correction for the mean scores obtained by the first, fourth, and eighth week groups are shown in Figure 1. Apart from *D*, *Hy*, and *Mf*, there was a consistent increase in the clinical scales from first, to fourth, to eighth week. A *t* test was used to evaluate the significance [2] of the differences between the means for the three groups on the MMPI main scales, the *Pd*, *Ma*, and *Pa* subscales, the ego-strength scale, the recidivism scale, and the pathology scale (Dahlstrom & Welsh, 1960). There were no significant differences between the first and fourth week groups. The eighth week group obtained higher scores than the fourth week group on *Pa* 1 (ideas of external influence, $p < .01$), *Ma* 1 (amorality, $p < .05$), and recidivism ($p < .01$). The eighth week group had a lower score than the fourth week group on *K* ($p < .05$). The eighth week group's scores were higher than the first week group on *F* ($p < 01$), *K* ($p < .01$), *Hs* ($p < .05$), *Pd* ($p < .05$), *Ma* ($p < .01$), *Pd* 4 (alienation, $p < .05$), *Pa* 1 ($p < .01$), *Ma* 1 ($p < .01$), *Ma* 4 (ego inflation, $p < .01$), recidivism ($p < .01$), and pathology ($p < .01$). The eighth week group's score on *K* ($p < .01$) was lower than the first week group.

There was no increase in scores on ego-strength, or any other evidence of beneficial psychological effects accruing from basic training. The change in the shape of the profiles suggests that aggressive, impulsive, and energetic features became slightly more prominent. The most cautious explanation of this change would be that the subjects were more willing to admit mild kinds of antisocial behavior in the eighth week, since they had finished their training, and perhaps also experienced some relief at completing

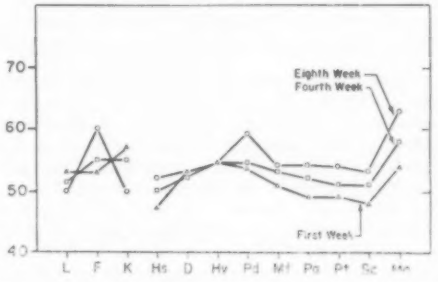


FIG. 1. Mean MMPI profiles without *K* correction for the first, fourth, and eighth week groups.

[1] An extended report of this study may be obtained without charge from Paul Ekman (University of California School of Medicine; San Francisco 22, California) or for a fee from the American Documentation Institute. Order Document No. 6875 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

[2] Two-tailed probability tests were either at .05 or .01 levels of confidence.

an arduous task. The changes on the subscales imply that more callous attitudes, a tendency to ignore the needs of others, and feelings of self-importance increase slightly during basic training. The recruits appear less prone to examine their own responsibility for conflicts, and more ready to react aggressively.

The fact that the eighth week group obtained a mean score on the recidivism scale close to that reported for men with multiple AWOLs, and a mean score on the pathology scale higher than that reported for neurotic veterans, raises some question as to the validity of these two scales.

REFERENCE

DAHLSTROM, W. G., & WELSH, G. S. *An MMPI handbook*. Minneapolis: Univer. Minnesota Press, 1960.

(Received January 31, 1961)

## STABILITY OF RORSCHACH PENETRATION AND BARRIER SCORES OVER TIME [1]

PAUL G. DASTON AND OWEN L. McCONNELL

*Duke University*

Fisher and Cleveland (1958) have evolved a Rorschach content scoring system they relate to body image. They categorize responses having vulnerable boundaries (e.g., mashed bug) as Penetration (P) and those with firm boundaries (e.g., cocoon) as Barrier (B). Across-groups, empirical evidence supports these as meaningful categorizations. Regarding their possible use in longitudinal studies, the present study assesses stability of B and P over time with normal subjects in a relatively constant environment.

The subjects were 20 males, hospitalized with long term physical disorders but screened for absence of psychiatric disorder. All were tested twice 2 months apart by the same experimenter, using Beck's instructions (Schneider, 1955). The 40 protocols, arranged randomly, were scored for B and P by two raters working independently.[2] To correct for correlation with response total, scores from the longer record of each subject were reduced in accordance with the ratio of his two response totals.

Both raters learned the scoring system via Fisher and Cleveland's text and had not conferred about scoring problems; so a check on interrater agreement seemed appropriate. Pearson *r*'s between raters were .84 for B and .61 for P. The P discrepancy required further investigation. When the two raters scored a new record together, they diverged on implied vs. explicit Penetration. After differences in interpretation were clarified, the raters scored another set of 48 records, with P agreement rising to an *r* of .79, certainly more acceptable.

Examining stability, test-retest *r*'s were for B and P respectively: Rater A, .90, .82; Rater B, .88, .79.[3] Differences between means for related samples were not significant for either rater on B or P scores. However, both raters found P scores to decrease on retest (almost at .05 level for Rater A). Since mild psychological changes probably occurred in the subjects between tests (e.g., improvement in health), P may be a more sensitive indicator of minimal change than B.

Overall, the Fisher-Cleveland scoring system seems sufficiently objective and stable under the conditions employed to be of value in assessing effects of various treatment conditions over time.

REFERENCES

FISHER, S., & CLEVELAND, S. E. *Body image and personality*. Princeton: Van Nostrand, 1958.

SCHNEIDER, B. H. The effect of varying time intervals on the reproduction and recall of Rorschach responses on retest. Unpublished doctoral dissertation, Michigan State University, 1955.

(Received March 6, 1961)

[1] An extended report of this study may be obtained without charge from Paul G. Daston (Department of Psychology, University of Maryland; College Park, Maryland) or for a fee from the American Documentation Institute. Order Document No. 6937 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

[2] We are grateful to William Groman, Duke University, for his assistance.

[3] Sidney Cleveland and Seymour Fisher were also kind enough to score these records for B and P. They minimized effects of differential response productivity on a card by card basis, after which they calculated *r*'s. Test-retest *r* for B was .93, for P, .90, which can be taken as independent evidence for the stability of the scoring system over time.

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 105

# A COMPARISON OF THE SHIPLEY-HARTFORD AS A POWER TEST WITH THE WAIS VERBAL SCALE [1]

H. J. WAHLER AND LUKE S. WATSON

*University Health Center, Ohio State University*

In many applied settings there is practical interest in efficient screening tests of intellectual functioning. One test frequently used has been the Shipley-Hartford. Correlations between scores on this test and the Wechsler-Bellevue have ranged from .69 to .79 as reported by Garfield and Fly (1948), Sines (1958), and Wright (1945). These authors used standard administration with 10-minute time limits per subtest and did not control time span between administrations of the two tests.

The purpose of this study was to determine the correlation between the WAIS Verbal IQ and the two S-H subtests, singly and in combination, when the S-H is given without time limits. Two samples were employed and the multiple regression weights for the S-H Vocabulary and Abstract Reasoning scores double cross-validated. Other relations investigated included WAIS Verbal IQ with S-H subtest scores combined with unit weights, the WAIS personally administered Vocabulary with the S-H self-administered multiple-choice Vocabulary, and the number of S-H multiple-choice vocabulary items attempted relative to WAIS Verbal IQ.

Two samples of white adults seeking help at a mental hygiene clinic and rehabilitation center were tested by two experimenters. The first sample consisted of 50 subjects—30 females and 20 males. The second sample was comprised of 23 females and 32 males. The WAIS Verbal Scale and S-H were administered during the same testing session in all cases.

The S-H total score (V + A) correlated .88 with WAIS Verbal IQ for Sample 1 and .87 for Sample 2. Multiple $R$ for the two S-H subtests with WAIS Verbal IQ was .88 for both samples. Whether S-H Vocabulary and Abstract scores were combined with unit or optimum linear weights, over 77% of the variance in WAIS Verbal scores was accounted for by variance in the combined Shipley scores; this was true for both samples. When $b$ weights derived from the multiple regression equation for the first set of data were applied to scores of the second sample, $R$ was .878; with $b$ weights reversed, $R$ was .877. The two methodologically different vocabulary tests correlate consistently high with $r$'s of .85 and .86 for Samples 1 and 2, respectively. It is also noteworthy that brighter people tended to attempt more S-H Vocabulary items than did duller subjects as reflected in the significant correlations between WAIS Verbal IQ and number of items attempted: $r = .54$ for Sample 1 and $r = .53$ for Sample 2.

When a double entry table based on six categories from the WAIS manual was used, 59% of the subjects were classified the same with S-H IQ estimates as with their WAIS Verbal IQs. Seventeen percent of the cases were overestimated and 20% underestimated by one category. Only 6% were missed by two categories. With a dichotomous classification of above or below average, S-H estimates correctly classified 87% of the subjects.

When IQ is a minor variable in a decision complex and only a broad categorization is required for the circumstances, the findings support the position that a Shipley estimate can be sufficient.

[1] An extended report of this study may be obtained without charge from H. J. Wahler (University Health Center; 320 West Tenth Avenue; Columbus 10, Ohio) or for a fee from the American Documentation Institute. Order Document No. 6938 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C. remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

## REFERENCES

GARFIELD, S. L., & FLY, W. A comparison of the Wechsler-Bellevue and Shipley-Hartford scales as measures of mental impairment. *J. consult. Psychol.*, 1948, **13**, 259–269.

SINES, L. K. Intelligence test correlates of Shipley-Hartford performance. *J. clin. Psychol.*, 1958, **14**, 399–404.

WRIGHT, E. M. Use of the Shipley-Hartford test in evaluating intellectual functioning of neuropsychiatric patients. *J. appl. Psychol.*, 1945, **30**, 45–51.

(Received February 17, 1961)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 106

# TEACHING RELIABLE RATING OF A PROCESS VARIABLE[1]

ROBERT RYDER

*University of Michigan*

Therapy process research has been hampered by poor rating reliability for the variables under study, particularly when ratings are based on short units of interaction. One difficulty may be that the variables have complex and subtle relations to concrete therapist behaviors. Thus one might logically analyze a concept into simpler, easier to measure components. Another approach is to compare rating conditions empirically to see which leads to the most reliable results; but reliability may be bought at the price of validity—in some cases reliability goes up as less information is given to the rater. Similarly, when ratings are averaged over raters the result may be a reliable score with what could be called a common denominator character. We may lose the fineness and texture of individual sensitivities. This paper describes an attempt to avoid these difficulties, using "therapist commitment" as the variable.

A total of 385 therapist statements were abstracted from interview transcripts and mounted on 3 × 5 cards. *One* experienced and sophisticated therapist rated each statement on "degree of commitment," using a four alternative scale. He rated 285 cards in the sequence in which they were originally spoken (higher information condition) and 100 cards in random sequence (lower information). Four graduate students in clinical psychology were then taught to postdict these (criterion) ratings by a flash-card procedure: The student read a statement, guessed what the criterion rating must be, then turned the card over—the correct criterion rating was on the back of the card. Training was continued for eight sessions, one per week, each session being followed by a cathartic discussion session, which was one technique used to maintain high motivation. Other techniques were in the social atmosphere that was maintained (the study was referred to as "the development of a process variable," not as a rating study, the students were called "participants," not raters, and a girl was hired to serve them coffee and otherwise emphasize their importance), and in wages paid—four dollars per hour. Participants were instructed to not mention particular statements in the discussion sessions.

Taus relating postdictions to criterion ratings were .94, .93, .92, and .71 for the final session, as compared to first session taus of .44, .35, .31, and .27. These taus refer to the relation between criterion ratings and participants' *first* guesses as to those ratings; i.e., each of these guesses was made when a participant was encountering a card he had never seen before. Our median tau of .925 seems favorable in relation to Williams' (1959) median interrater product-moment reliability, also for commitment, of .68 (she trained her raters by having discussions with them), and with the retest reliability (tau, after a 3 month period) of our criterion rater, which was also .68. We conclude that one may find participants who by using this technique can achieve a great increment in consistency in their use of the words "therapist commitment." Note that the concept is as stable over experiments as we choose to make it, as it is entirely defined by a concrete set of criterion ratings.

There was no relation of postdiction accuracy to either level of information or to (rated) level of participants' motivation.

[1] An extended report of this study may be obtained without charge from Robert Ryder (Department of Psychology, University of Michigan; Ann Arbor, Michigan) or for a fee from the American Documentation Institute. Order Document No. 6935 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $2.00 for microfilm or $3.75 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

## REFERENCE

WILLIAMS, JOAN V. The influence of therapist commitment on progress in psychotherapy. Unpublished doctoral dissertation, University of Michigan, 1959.

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 107

# DEGREE OF AGREEMENT AMONG THREE "INTERPERSONAL SYSTEM" MEASURES[1]

MALCOLM D. GYNTHER

*Washington University Medical School*

Leary (1957) states that public interpersonal behavior (Level I) is measured by the MMPI and sociometric ratings and that private self-description (Level II) is measured by the Interpersonal Check List (ICL). He has found significant positive relations between MMPI data and sociometric ratings of behavior of patients in group psychotherapy, but has not compared self-ratings with the other measures. The aim of this study is to determine the agreement among interpersonal circle classifications of nonpsychiatric subjects obtained by these three assessment techniques.

Sixty-seven male subjects and 33 female subjects, graduate psychology students ($N = 8$), five groups of clinical pastoral trainees ($N = 30$), and four classes of college extension students ($N = 62$), participated. Each group was small enough to allow considerable interaction. Both sexes were used as previous work had shown no significant differences referable to sex between octant distributions of MMPIs and ICLs for nonpsychiatric subjects. Subjects were given MMPIs and ICLs about 1 month after their classes began, with instructions to rate self and the other members of their group on the ICL. Groups larger than eight were divided so that no more than eight subjects would be in any one rating group. These groups ranged in size from four to eight, with a median of six members.

MMPI standard scores were converted into Dominance (Dom) and Love (Lov) scores by means of the following formulae: $\text{Dom} = (Ma - D) + (Hs - Pt)$ and $\text{Lov} = (K - F) + (Hy - Sc)$. These indices determine the octant rating. Similar procedures yielded octant ratings of ICL data. Sociometric ratings were obtained by adding the Dom and Lov scores given to a subject by the other subjects in his group, dividing the totals by the number of "others," and determining an octant rating for the resulting values. Agreement between different measures was considered to exist if the different ratings fell in the same octant, a more rigorous criterion than Leary uses.

Most of the ratings fell in what Leary has described as the "healthy" octants (i.e., 2, 1, 8, 7). Agreement between ratings may be spuriously high due to these constricted distributions. The percentages of "hits" between classifications derived from MMPI and sociometric, self-description and sociometric, and MMPI and self-description were 26, 42, and 21, respectively. Chi square analysis (based on Octants 1, 8, and "other") of MMPI and sociometric data showed that agreement was not significantly different from chance. However, agreement between self-descriptions and sociometric ratings, using Octants 1, 2, and other categories, was highly significant ($\chi^2 = 13.89$, $df = 4$, $p < .01$). Analysis (based on Octants 1, 8, and other) of MMPI and self-rating data also showed a significant degree of agreement ($\chi^2 = 9.52$, $df = 4$, $p < .05$).

Operating within the framework of the Interpersonal System, one would predict more agreement between ratings presumably tapping the same level (MMPI and sociometric) than ratings supposedly tapping different levels (MMPI and self-description and sociometric and self-description). Our contradictory findings cast doubt on Leary's assertion that these tests measure specified, distinct levels of behavior. These results also indicate that the positive MMPI-sociometric relationship found with psychiatric subjects does not hold with nonpsychiatric subjects.

[1] An extended report of this study may be obtained without charge from Malcolm D. Gynther (1420 Grattan Street; St. Louis 4, Missouri) or for a fee from the American Documentation Institute. Order Document No. 6936 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

## REFERENCE

LEARY, T. *Interpersonal diagnosis of personality*. New York: Ronald, 1957.

(Received April 5, 1961)

*Journal of Consulting Psychology*
1962, Vol. 26, No. 1, 108

# ANXIETY INDICES IN HUMAN FIGURE DRAWINGS: A REPLICATION AND EXTENSION[1]

ROBERT E. MOGAR

*Pennsylvania State University*

Hoyt and Baron (1959) tested the validity of 12 presumed anxiety indicators on the Draw-A-Person Test (DAP) using the Manifest Anxiety (*MA*) scale as an external criterion. It was found that Placement (upper left hand corner) and Size (less than 4½ inches) were displayed by a significantly greater proportion of anxious than nonanxious female psychiatric patients.

The present study replicated the Hoyt-Baron study with a male psychiatric sample. In addition, the validity of the potential anxiety indicators on the DAP was tested against a second anxiety criterion; the Rorschach Content Test (RCT) developed by Elizur (1949). The relationship of each DAP index to intelligence was also estimated.

The drawings of 123 subjects were scored for presence or absence of each drawing characteristic following the Hoyt-Baron scoring system. Scoring reliabilities ranged from .84 to 1.00. Point biserial correlations were computed between each DAP anxiety index and *MA* scale scores, RCT scores, and intelligence (as measured by the Verbal WAIS). In addition, chi-square analyses of differences in proportions of anxious and nonanxious subjects displaying each index were computed using both the *MA* scale and RCT as bases for forming groups.

There was substantial agreement between the results of the Hoyt-Baron study and the present replication in that little relationship was found between the DAP anxiety indices and the *MA* scale. Only Relative Size of Head to Figure (large) was significantly related to *MA* scale scores ($r = .17$). Contrary to Hoyt and Baron, neither Placement nor Size significantly differentiated extreme *MA* groups. These disparate results seemed primarily a function of the sex of the samples employed.

The expectation that the RCT would provide a more appropriate criterion of clinical anxiety than the *MA* scale was supported by the finding that size ($r = .26$), Relative Size of Head to Figure ($r = .24$), Omissions ($r = .18$), and Hair Shading ($r = .14$) were significantly related to RCT scores. A negative relationship was found between Reinforcement and RCT scores ($r = -.31$). Intelligence was significantly related to Erasing ($r = .21$) and Relative Size of Head to Figure ($r = -.17$).

These results indicated that the *MA* scale does not measure anxiety along the same continuum as the RCT and DAP. The lack of relationship between *MA* and RCT scores ($r = .07$) supports this conclusion. The present study also suggests that indirect measures of anxiety such as the RCT which can be reliably quantified may be less convenient than paper-and-pencil tests but perhaps more conceptually and clinically useful.

[1] An extended report of this study may be obtained without charge from Robert E. Mogar (Mt. Zion Psychiatric Clinic; 2255 Post Street; San Francisco 15, California) or for a fee from the American Documentation Institute. Order Document No. 6934 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

## REFERENCES

ELIZUR, A. Content analysis of the Rorschach with regard to anxiety and hostility. *Rorschach res. Exch.*, 1949, **13**, 247–283.

HOYT, T. F., & BARON, M. R. Anxiety indices in same-sex drawings of psychiatric patients with high and low MAS scores. *J. consult. Psychol.*, 1959, **23**, 448–452.

(Received April 17, 1961)

**From four years of intensive research . . . new and valuable information on**

# THE PSYCHOTHERAPY RELATIONSHIP

**By William U. Snyder,**
**The Pennsylvania State University**
**With the collaboration of B. June Snyder**

A four-year research project has produced this valuable book, which offers an intensive study of the psychotherapy relationship. The author presents extensive case studies, demonstrating the methods employed and focusing upon the dynamic relationship between therapist and client. Dr. Snyder's eclectic approach places strong emphasis on the interpersonal relationship formed during counseling and therapy, and demonstrates techniques and skills that are unusually flexible and effective. *The Psychotherapy Relationship* is a basic text for courses in psychotherapy and a supplementary text and reference for courses in Personality and Abnormal Psychology. Some of the features that will make you want to have a copy for your personal use are:

- Correlated research and questionnaire data for 550 recorded interviews.
- Discussion of the use of relationship dreams in psychotherapy.
- Illustration and analysis of therapeutic process.
- Many examples of positive and negative transference and of the therapist's countertransference.
- Thorough examination of one case study and selections from 19 others.

**EXAMINE THIS INFORMATIVE NEW STUDY FOR TEN DAYS WITHOUT OBLIGATION.**

---

The Macmillan Company, 60 Fifth Avenue, New York 11, N. Y. Box JC-1

Please send me a copy of Snyder: THE PSYCHOTHERAPY RELATIONSHIP @ $7.50
I may return it in ten days without cost or obligation. Otherwise bill me. (Save! Send payment with your order and we pay for delivery.)

☐ Payment enclosed ☐ Bill me

NAME.....

STREET.....

CITY.....ZONE...STATE.....

## COUNSELING FOR PERSONAL ADJUSTMENT

In Schools and Colleges

**Fred McKinney**

This text for the first course in counseling concentrates on certain basic principles and techniques, with special emphasis on the nature of personality structure and dynamics. Sixty case studies illustrate counseling in an educational context.

*584 pages* *1958* *$6.75*

## CLIENT-CENTERED THERAPY

**Carl R. Rogers**

A mature presentation of the non-directive and related points of view in counseling and therapy, this volume orients the professional public to client-centered therapy and details its applications and implications to several fields, including work with children and groups.

*560 pages* *1951* *$5.75*

Houghton Mifflin

BOSTON · NEW YORK · ATLANTA

GENEVA, ILL. · DALLAS · PALO ALTO

*1 One of a series*

# Tall Oaks From Little Acorns Grow

*David Everett*

Through the ages, man has held a tiny acorn in his hand and envisioned the future giant of the forest.

What laws of nature must operate to assure the transition from insignificance to full growth?

*Placement* in fertile soil; *protection* of the life essence; *nourishment* from natural elements; *time* to develop into the type of tree nature intended — these combine to form the *right environment.*

So life's power in individuals comes to its full realization through proper conditions working to nurture innate potentials.

Aware of latent possibilities in each person entrusted to its care, Devereux Schools — now celebrating its 50th Anniversary — has grown from a very small beginning to serve today more than a thousand children and young adults in schools, communities, and camps in four states. Through the past half century, it has provided the right environment for thousands of mentally retarded and emotionally disturbed children who have benefitted from it and returned to their families and communities.

Thus, from the acorn of an idea has grown a stalwart tree . . . . THE DEVEREUX FOUNDATION.

•

Your inquiries are invited. For further information and literature, East Coast residents may contact Mr. Charles J. Fowler, Director of Admissions, Devereux Schools, Devon, Pa. Those in the West Coast area, Mr. Keith A. Seaton, Registrar, Devereux Schools in California, Santa Barbara, Calif. Southwestern residents, Mr. Richard Danko, Executive Administrator, Devereux Schools in Texas, Victoria, Texas.

** Since the founding of Devereux Schools in 1912, a philosophy of Special Education and therapeutic techniques of rehabilitation and care have been developed for the sole reason of insuring to the greatest possible extent the future happiness and well-being of the retarded or emotionally disturbed child. Many of these techniques originated at Devereux Schools. This series is launched during the Schools' 50th Anniversary in order that you may better understand the methods and goals of Devereux Schools.*

## The Devereux Foundation

Helena T. Devereux, *Founder and Consultant* • Edward L. French, Ph.D., *President and Director*
A nonprofit organization . . . Devon, Penna. • Santa Barbara, Calif. • Victoria, Texas
SCHOOLS • COMMUNITIES • CAMPS • TRAINING • RESEARCH

*Half a Century of Service to Children*

"The best in tests. . . ."

Could you use a projective test eliciting a rich variety of unguarded responses from deep layers of the personality? Easily and quickly administered? Giving results of demonstrated reliability and validity? For children as well as adults?

The

# KAHN TEST OF SYMBOL ARRANGEMENT

uses simple plastic objects arranged by the examinee in various sequences according to several different sets of easily-followed instructions. An excellent supplement to or substitute for the TAT and Rorschach, combining the principal advantages of both.

•

Obtain information and/or order (complete set $25.00) from

PSYCHOLOGICAL TEST SPECIALISTS

Box 1441
Missoula, Montana

"The best in tests. . . ."